1. Oscillations in Mechanical Systems
    1. Simple Harmonic Oscillator
    2. Module 2.xhtml
    3. Module 3.xhtml
    4. Module 4.xhtml
    5. Module 5.xhtml
    6. Module 6.xhtml
2. Module 7.xhtml
3. Mechanical Waves
    1. Module 8.xhtml
    2. Waves
    3. Module 10.xhtml
    4. Energy in a mechanical wave
    5. Module 13.xhtml
    6. Fourier Series
4. Div, Grad, Curl
    1. Scalar.xhtml
    2. Vector Fields
    3. Vector multiplication.xhtml
    4. Gradient.xhtml
    5. Gauss' Theorem
    6. Module 17.xhtml
5. Electromagnetism Review
    1. Module 18.xhtml
    2. Module 19.xhtml
    3. Module 20.xhtml
6. Electromagnetic Waves
    1. Module 21.xhtml
    2. module 22.xhtml
    3. Spherical Waves
    4. module 24.xhtml

5. Module 25.xhtml
6. Module 26.xhtml
7. Module 27.xhtml
8. Module 28.xhtml

7. Optics
    1. Module 29.xhtml
    2. Module 30.xhtml
    3. Electromagnetism at an Interface
    4. Snell's Law
    5. The Fresnel Equations
    6. Module 34.xhtml
    7. Evanescent Wave
    8. Phase Changes
    9. Module 35.xhtml
    10. Polarization
8. Geometric Optics
    1. Refraction at a Spherical Interface
    2. Thin Lens
    3. Mirrors
    4. Eyeglasses
9. Interference
    1. Two Source Interference
    2. Thin Film Interference
    3. DoublePlateInterference.xhtml
    4. NewtonsRings.xhtml
    5. Multi Source Inteference
10. Diffraction
    1. SingleSlitDiffraction.xhtml
    2. DoubleSlitDiffraction.xhtml
    3. DiffractionGrating.xhtml
    4. Diffraction from a Rectangular Aperture
    5. Babinet's Principle

    6. Diffraction from a Circular Aperture
    7. Diffraction from an Array of Apertures
11. Fourier Optics
    1. Fourier Optics
    2. Dirac Delta Function
    3. The Convolution Theorem and Diffraction
    4. An Array of Apertures

Simple Harmonic Oscillator

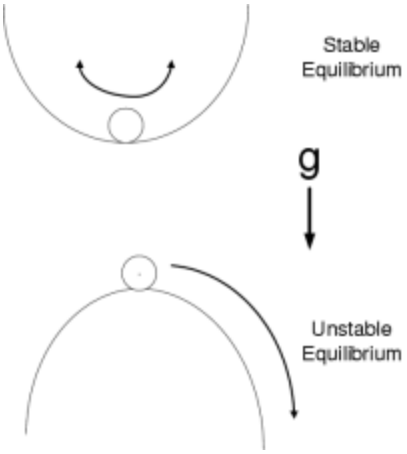
The Simple Harmonic Oscillator

Simple Harmonic Motion

For SHM to occur we require stable equilibrium, about a point. For example, at the origin we could have:

$$\sum \vec{F}(0) = 0,$$

which would describe a system in equilibrium. This however is not necessarily stable equilibrium.

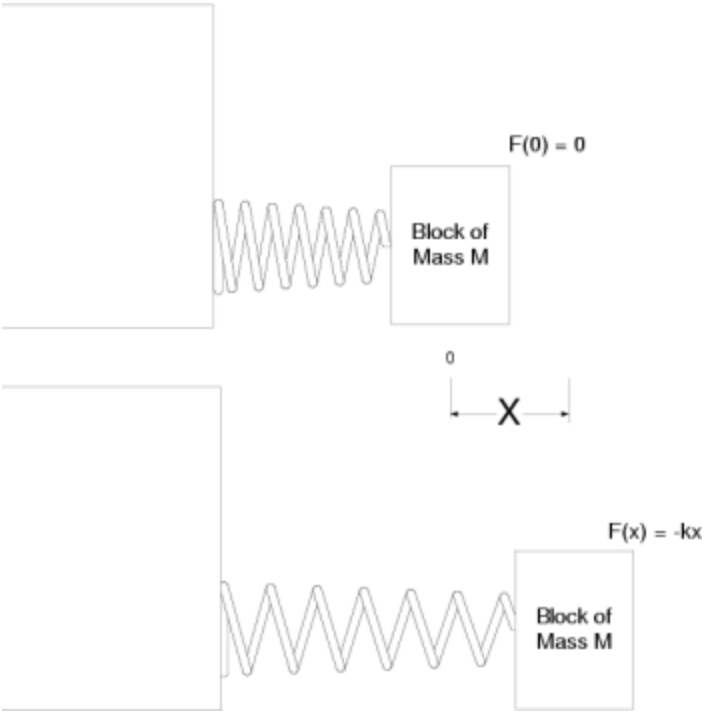


A simple cartoon of stable and unstable equilibrium. The lower part of the figure shows the case of unstable equilibrium. The upper part shows the case of stable equilibrium. These

situations often occur in mechanical systems.

The lower part of the figure shows the case of unstable equilibrium. The upper part shows the case of stable equilibrium. These situations often occur in mechanical systems.

For example, consider a mass attached to a spring:



In general, in a case of stable equilibrium we can write the force as a polynomial expansion:

$$F(x) = -(k_1 x + k_2 x^2 + k_3 x^3 + \ldots)$$

where the k_i are positive constants. There is always a region of x small enough that we can write:

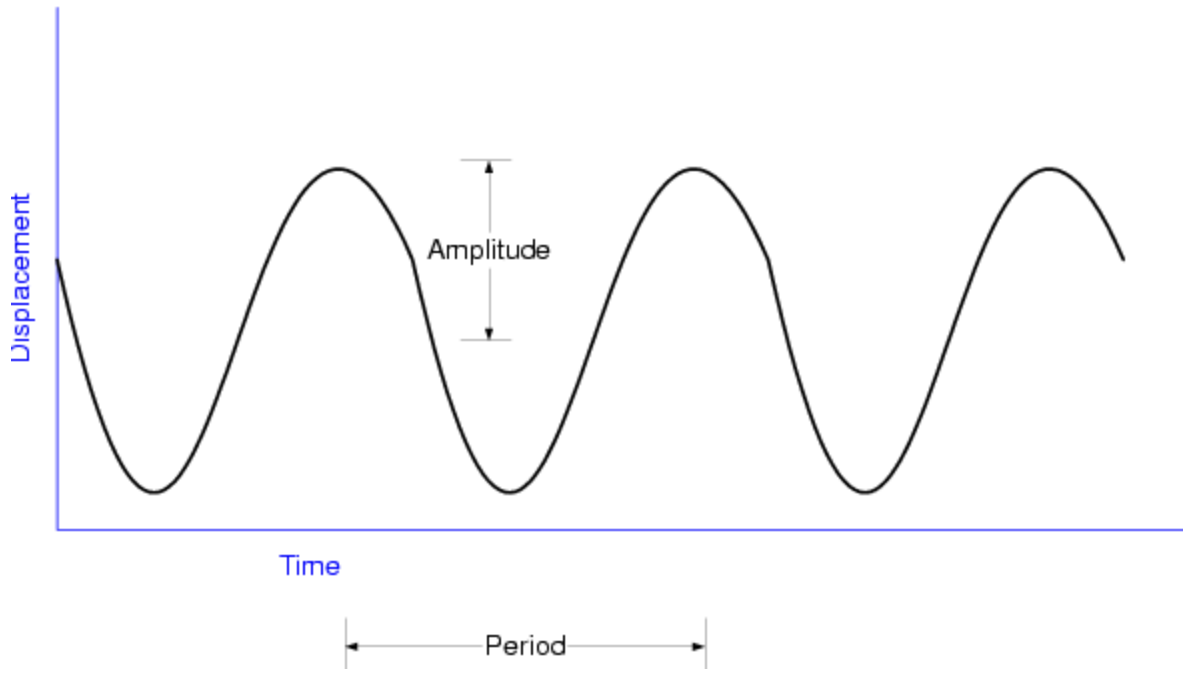
$$F = -kx$$

$$F = -kx$$
$$ma = -kx$$
$$m\ddot{x} = -kx$$
$$\ddot{x} + \tfrac{k}{m}x = 0$$

This is satisfied by an equation of the form

$$x = A\sin\left(\omega t + \varphi_0\right)$$

where A and φ_0 are constants that are determined by the initial conditions. Draw a diagram of a sinusoid and mark on it the period T and Amplitude A



φ_0 Is an arbitrary phase which shifts the sinusoid.This is also satisfied by an equation of the form

$$x = A\sin(\omega t) + B\cos(\omega t)$$

Lets show this:

$$\begin{aligned} x &= A\sin(\omega t) + B\cos(\omega t) \\ \dot{x} &= \omega(A\cos(\omega t) - B\sin(\omega t)) \\ \ddot{x} &= -\omega^2(A\sin(\omega t) + B\cos(\omega t)) \\ \ddot{x} &= -\omega^2 x \end{aligned}$$

Again there are two constants determined by the initial conditions A and B
The equation can be rewritten

$$\ddot{x} + \omega^2 x = 0$$

Thus if

$$\omega^2 = \frac{k}{m}$$

then the equation is identical to the SHM equation.

So another way to write the equation of Simple Harmonic Motion is

$$\ddot{x} + \omega^2 x = 0$$

or

$$\ddot{x} = -\omega^2 x$$

It is also important to remember the relationships between freqency, angular frequency and period:

$$\begin{aligned} \omega &= 2\pi\nu \\ T &= \tfrac{2\pi}{\omega} \\ \nu &= \tfrac{1}{T} \end{aligned}$$

Another solution to the SHM equation is

$$\widetilde{x} = A\cos\left(\omega t + \varphi_0\right) + iA\sin\left(\omega t + \varphi_0\right)$$

Recall Taylor's expansions of sine and cosine

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!}\dots$$

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!}\dots$$

Then

$$\begin{aligned}\cos\theta + i\sin\theta &= 1 + i\theta - \tfrac{\theta^2}{2!} - i\tfrac{\theta^3}{3!} + \tfrac{\theta^4}{4!}\dots \\ &= 1 + i\theta + \tfrac{(i\theta)^2}{2!} + \tfrac{(i\theta)^3}{3!} + \tfrac{(i\theta)^4}{4!}\dots \\ &= e^{i\theta}\end{aligned}$$

(an alternative way to show this is the following)

$$\begin{aligned}z &\equiv \cos\theta + i\sin\theta \\ \mathrm{d}z &= (-\sin\theta + i\cos\theta)\mathrm{d}\theta = iz\mathrm{d}\theta \\ \int \tfrac{\mathrm{d}z}{z} &= \int i\mathrm{d}\theta \\ \ln z &= i\theta \\ z &= e^{i\theta}\end{aligned}$$

Thus we can write

$$\widetilde{x} = A\cos\left(\omega t + \varphi_0\right) + iA\sin\left(\omega t + \varphi_0\right)$$

as

$$\widetilde{x} = Ae^{i(\omega t + \varphi_0)}$$

$$\begin{aligned}\widetilde{x} &= Ae^{i(\omega t + \varphi_0)} \\ \dot{\widetilde{x}} &= i\omega Ae^{i(\omega t + \varphi_0)} \\ \ddot{\widetilde{x}} &= (i\omega)^2 Ae^{i(\omega t + \varphi_0)} = -\omega^2\widetilde{x}\end{aligned}$$

Note: We will use the complex representation a lot, so you need to become familiar with it. It is used a lot in Optics, Classical and Quantum Mechanics and Electrical Engineering so it is a good thing to know.

Now for physical systems we are interested in just the real part so

$$x = Re\left[Ae^{i(\omega t+\varphi_0)}\right]$$

This will be implicitly understood. In physics we just write

$$x = Ae^{i(\omega t+\varphi_0)}$$

One thing that will seem to be confusing is that there are all these different solutions. They are all just different forms of the same thing. Which form is used in a particular circumstance is simply a matter of convenience. Some forms lend themselves to to solutions of certain problems more easily than others. Also the most convenient form can depend upon the initial conditions. For example if x is at its maximum displacement at time $t = 0$ then a cos form may be the most convenient. As a general rule I like using the complex representation because natural logarithms are so easy to work with. For example

$$\frac{\mathbb{d}e^x}{\mathbb{d}x} = e^x$$

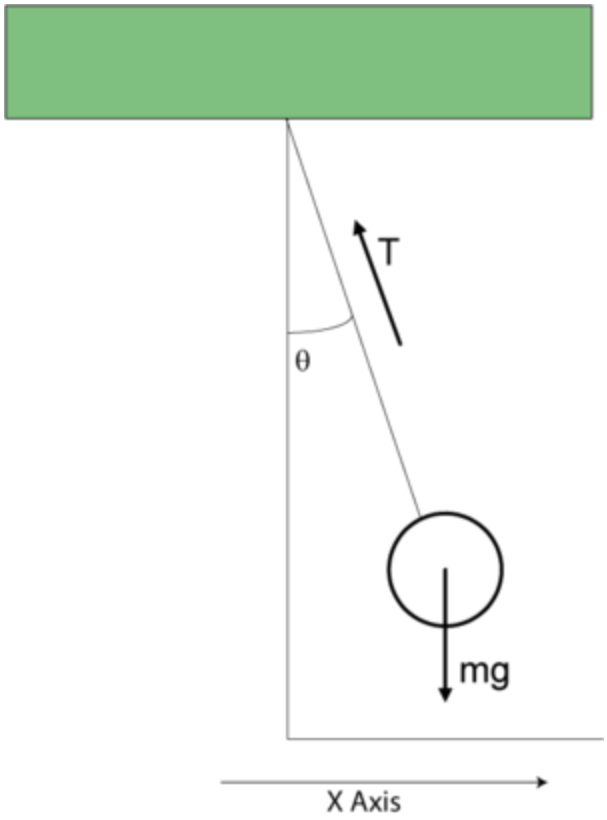
$$\frac{\mathbb{d}e^{ax}}{\mathbb{d}x} = ae^{ax}$$

$$\int e^{ax}\mathbb{d}x = \frac{1}{a}e^{ax}$$

which is all pretty simple to remember

Module 2.xhtml

The Simple Pendulum



A simple pendulum.

Shown is a simple pendulum which has a mass m that is displaced by an angle θ. There is tension ($\vec{T}$) in the string which acts from the mass to the anchor point. The weight of the mass is $m\vec{g}$ and the tension in the string is $T = mg\cos\theta$. There is a tangential restoring force $= -mg\sin\theta$. If we approximate that θ is small (we have to make this approximation or else we can not solve the problem analytically) then $\sin\theta \approx \theta$ and $x = l\theta$. (note that $\sin\theta$ is only approximately equal to $\frac{x}{l}$ because x is the distance along the x axis) so that we can write:

$$
\begin{aligned}
F = ma = m\ddot{x} \\
= -mg\sin\theta \\
\approx -mg\theta \\
\approx -mg\tfrac{x}{l}
\end{aligned}
$$

or

$$\ddot{x} + \frac{g}{l}x = 0$$

(Note that We should immediately recongnize that this is the equation for simple harmonic motion (SHM) with

$$\omega = \sqrt{\frac{g}{l}}.$$

We could take another approach and use angular momentum to solve the problem. Recall that:

$$L = I\omega = I\dot{\theta}$$

$$I = ml^2.$$

Also recall that the torque is the time derivative of the angular momentum so that:

$$\vec{\tau} = \vec{r} \times \vec{F} = \tfrac{\partial \vec{L}}{\partial t}$$

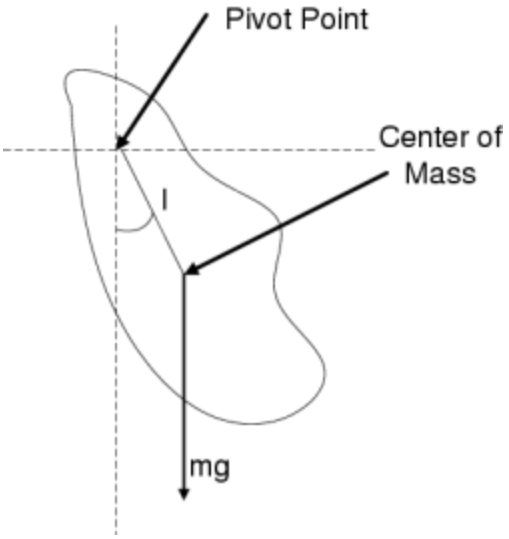
$$-lmg\theta = I\ddot{\theta}$$

$$\ddot{\theta} + \frac{g}{l}\theta = 0$$

Again we would recognize that this is simple harmonic motion with

$$\omega = \sqrt{\frac{g}{l}}.$$

The Compound Pendulum



A compound pendulum.

The compound pendulum is another interesting example of a pendulum that undergoes simple harmonic motion. For an extended body then one uses the center of mass and the moment of inertia. Use the center of mass, the moment of inertia and the Torque (angular force) $\vec{\tau} = \vec{r} \times \vec{F}$

$$\tau = r \times F$$

$$I\ddot{\theta} = -lmg \sin\theta \approx -lmg\theta$$

$$\ddot{\theta} + \tfrac{lmg}{I}\theta = 0$$

So again we get SHM now with

$$\omega^2 = \frac{lmg}{I}$$

One sees that this formalism can be applied to the simple pendulum (ignore the string and one can consider the ball a point mass). The moment of

inertia is ml^2. So we get

$$\omega^2 = \frac{lmg}{ml^2} = \frac{g}{l}$$

which is just what we got before for the simple pendulum. We could write the equation of motion for a simple pendulum as:

$$\theta = Ae^{i(\omega t + \varphi_0)}$$

where φ_0 is determined by initial conditions.

A discussion of the Pendulum and Simple Harmonic Oscillator can be found at

http://monet.physik.unibas.ch/~elmer/pendulum/index.html

Module 3.xhtml

Same Frequency, different phase

One of the most important concepts we encounter in vibrations and waves is the principle of superposition. Lets look at a couple of cases starting with adding two motions with the same frequency but different phases. It is easiest to calculate this if you use complex notation

$$x_1 = A_1 e^{i(\omega t + \alpha_1)}$$
$$x_2 = A_2 e^{i(\omega t + \alpha_2)}$$

$$x = x_1 + x_2 = A_1 e^{i(\omega t + \alpha_1)} + A_2 e^{i(\omega t + \alpha_2)}$$
$$x = e^{i(\omega t + \alpha_1)} \left[A_1 + A_2 e^{i(\alpha_2 - \alpha_1)} \right]$$

This comes up all the time in real life: For example noise canceling headphones use this technique. In headphones there is a membrane vibrating with the frequency of the sound you are listening two. In a noise canceling headphone there is also a microphone "listening" to the noice coming from outside the headphone. This oscillation is inverted and then added to membrane producing the sound you listen to. The net result is a signal that contains the desired sound and subtracts the noise resulting in quieter operation.

Different Frequency

One can also consider the case of two oscillations with the same phase but different frequencies:

$$x_1 = A_1 e^{i(\omega_1 t)}$$
$$x_2 = A_2 e^{i(\omega_2 t)}$$

$$x = x_1 + x_2 = A_1 e^{i(\omega_1 t)} + A_2 e^{i(\omega_2 t)}$$
$$x = e^{i(\omega_1 t)} \left[A_1 + A_2 e^{i(\omega_2 - \omega_1) t} \right]$$

In an acoustical system, this gives beats, which is more easily seen if we take the case where $A_1 = A_2 \equiv A$, then:

$$\begin{aligned} x = x_1 + x_2 &= Ae^{i(\omega_1 t)} + Ae^{i(\omega_2 t)} \\ &= Ae^{i\left(\frac{\omega_1+\omega_2}{2}+\frac{\omega_1-\omega_2}{2}\right)t} + Ae^{i\left(\frac{\omega_1+\omega_2}{2}-\frac{\omega_1-\omega_2}{2}\right)t} \\ &= Ae^{i\left(\frac{\omega_1+\omega_2}{2}\right)t}\left[e^{i\left(\frac{\omega_1-\omega_2}{2}\right)t} + e^{-i\left(\frac{\omega_1-\omega_2}{2}\right)t}\right] \\ &= 2Ae^{i\left(\frac{\omega_1+\omega_2}{2}\right)t}\cos\left[\left(\frac{\omega_1-\omega_2}{2}\right)t\right] \end{aligned}$$

Where the last step used

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}$$

So in an acoustical system we will get a dominant sound that has the average of the two frequencies and and envelope of amplitude that slowly oscillates. This will be looked at more closes in the context of mechanical waves.

Module 4.xhtml

Energy in SHO

Recall that the total energy of a system is:

$$E = KE + PE = K + U$$

We also know that the kinetic energy is

$$K = \frac{1}{2}m\mathbf{v}^2$$

But what is U? For a conservative Force ($\oint \vec{F}d\vec{x} = 0$) - eg. gravity, electrical... (no friction) we know that the work done by an external force is stored as U. For the case of a mass on a spring, the external force is opposite the spring Force (That is it has the opposite sign from the spring force).:

$$F_{ext} = kx$$

(i.e. This is the force you use to pull the mass and stretch the spring before letting go and making it oscillate.)

Thus

$$U = \int_0^x kx\mathbb{d}x = \frac{1}{2}kx^2$$

This gives:

$$E = \tfrac{1}{2}m\mathbf{v}^2 + \tfrac{1}{2}kx^2$$
$$= \tfrac{1}{2}m\left(\frac{\mathbb{d}x}{\mathbb{d}t}\right)^2 + \tfrac{1}{2}kx^2$$

It is important to realize that any system that is represented by either of these two equations below represents oscillating system

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + kx = 0$$

$$\frac{1}{2}m\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \frac{1}{2}kx^2 = E$$

To calculate the energy in the system it is helpful to take advantage of the fact that we can calculate the energy at any point in x. For example in the case of the simple harmonic oscillator we have that:

$$x = Ae^{i(\omega t + \alpha)}$$

We can choose t such that

$$x = A$$

Now remember that when I write

$$x = Ae^{i(\omega t + \alpha)}$$

I "really" (pun intended) mean

$$x = Re\left[Ae^{i(\omega t + \alpha)}\right]$$

Likewise then

$$\dot{x} = Re\left[i\omega Ae^{i(\omega t + \alpha)}\right]$$

At the point in time where $x = A$ this gives us

$$\dot{x} = Re\left[i\omega A\right] = 0$$

Thus at that point in time we have $\dot{x} = 0$. We can now substitute that and $x = A$ into

$$E = \frac{1}{2}m\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \frac{1}{2}kx^2$$

we obtain

$$E = \frac{1}{2}kA^2$$

This is an important point. The energy in the oscillator is proportional to the amplitude squared!

Module 5.xhtml

Damped Oscillations

Consider a simple harmonic oscillator that has friction, then the equations of motion must be changed with the addition of a friction term. So we write

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -kx - b\frac{\mathrm{d}x}{\mathrm{d}t}$$

where $b\frac{\mathrm{d}x}{\mathrm{d}t}$ is the friction term. Rearranging we obtain:

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + b\frac{\mathrm{d}x}{\mathrm{d}t} + kx = 0$$

or

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \gamma\frac{\mathrm{d}x}{\mathrm{d}t} + \omega_0^2 x = 0$$

Where $\gamma = \frac{b}{m}$ and $\omega_0^2 = \frac{k}{m}$ Assume a solution of form

$$x = Ae^{i(pt+\alpha)}$$

substitute into equation and get

$$(-p^2 + ip\gamma + \omega_0^2)Ae^{i(pt+\alpha)} = 0$$

so

$$-p^2 + ip\gamma + \omega_0^2 = 0$$

p must have real and imaginary parts, so rewrite: $p = \omega + is$

$$p^2 = \omega^2 + 2i\omega s - s^2$$

So the equation

$$-p^2 + ip\gamma + \omega_0^2 = 0$$

becomes upon substitution:

$$-\omega^2 - 2i\omega s + s^2 + i\omega\gamma - s\gamma + \omega_0^2 = 0$$

This equation implies that the real and imaginary parts are each zero.Separate the real and imaginary partsImaginary parts give:

$$-2\omega s + \omega\gamma = 0$$
$$s = \tfrac{\gamma}{2}$$

From Real parts get

$$-\omega^2 + s^2 - s\gamma + \omega_0^2 = 0$$

or

$$-\omega^2 + \frac{\gamma}{4}^2 - \frac{\gamma}{2}\gamma + \omega_0^2 = 0$$

$$-\omega^2 - \frac{\gamma}{4}^2 + \omega_0^2 = 0$$

Which rearranges to

$$\omega^2 = \omega_0^2 - \frac{\gamma}{4}^2$$

Thus the solution becomes

$$x = Ae^{-\gamma t/2}e^{i(\omega t + \alpha)}$$

where

$$\omega = \sqrt{\omega_0^2 - \frac{\gamma}{4}^2}$$

Note that this has assumed a frictional damping force. For a more complicated damping force, the result would be different.

Module 6.xhtml

Lets consider a case with a driven or forced oscillator. We now have

$$m\ddot{x} + kx = F_0 e^{i\omega t}$$

or

$$\ddot{x} + \omega_0^2 x = \frac{F_0}{m} e^{i\omega t}.$$

Try

$$x = Ae^{i(\omega t + \alpha)}$$

then

$$\dot{x} = i\omega A e^{i(\omega t + \alpha)}$$

and

$$\ddot{x} = -\omega^2 A e^{i(\omega t + \alpha)}.$$

So now get

$$-\omega^2 A e^{i(\omega t + \alpha)} + \omega_0^2 A e^{i(\omega t + \alpha)} = \frac{F_0}{m} e^{i\omega t}$$

$$(\omega_0^2 - \omega^2) A e^{i(\omega t + \alpha)} = \frac{F_0}{m} e^{i\omega t}$$

$$(\omega_0^2 - \omega^2) A = \frac{F_0}{m} e^{-i\alpha}$$

$$(\omega_0^2 - \omega^2) A = \frac{F_0}{m} \cos\alpha - i\frac{F_0}{m} \sin\alpha$$

Next you spearate the Real and Imaginary parts. The Imaginary part gives

$$0 = -\frac{F_0}{m}\sin\alpha$$

so $\alpha = 0, \pi \ldots$ implies $\cos\alpha = \pm 1$ The Real part gives

$$(\omega_0^2 - \omega^2)A = \frac{F_0}{m}\cos\alpha$$

$$(\omega_0^2 - \omega^2)A = \pm\frac{F_0}{m}$$

$$A = \frac{\pm\frac{F_0}{m}}{(\omega_0^2 - \omega^2)}$$

So a solution is

$$x = \frac{\frac{F_0}{m}}{(\omega_0^2 - \omega^2)}e^{i(\omega t + \alpha)}$$

where $\alpha = 0, \pi \ldots$

This is an extremely important result, this is the phenomenum of resonance. When you drive an oscillator at its resonant frequency then the amplitude of the oscillation will become huge. In the equation above, it becomes infinite, but in practice there will be some damping that prevents that. You have known this since your childhood, this is how you swing on a swing. If you live in a snowy climate, you know (or at least should know) that a trick to get your car out of a snow bank is is to rock it back and forth - if you get the frequency right you will make the car oscillate with a large amplitude and dislodge it. The electrical analogue is used to tune a radio.

A driven damped oscillator will be given as a homework problem.

Module 7.xhtml

Partial Derivatives

A Partial derivative is defined as the derivative of the function w.r.t. one of the variables while holding the others constant

$$\frac{\partial f}{\partial x} = \lim_{\Delta x \to 0} \frac{f\left(x + \Delta x, t\right) - f\left(x, t\right)}{\Delta x}$$

$$\frac{\partial f}{\partial t} = \lim_{\Delta t \to 0} \frac{f\left(x, t + \Delta t\right) - f\left(x, t\right)}{\Delta t}$$

Some examples:

$$f(x,t) = 3x^2 + xt^2$$

$$\frac{\partial f}{\partial x} = 6x + t^2$$

$$\frac{\partial f}{\partial t} = 2xt$$

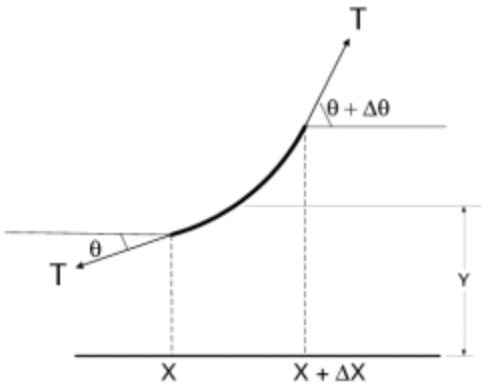
$$\frac{\partial^2 f}{\partial x^2} = 6$$

$$\frac{\partial^2 f}{\partial t^2} = 2x$$

$$\frac{\partial^2 f}{\partial x \partial t} = \frac{\partial^2 f}{\partial t \partial x} = 2t$$

Module 8.xhtml

Vibrations on a String



Consider the forces on a short fragment of string

$$F_y = T\sin\left(\theta + \Delta\theta\right) - T\sin\theta$$

$$F_x = T\cos\left(\theta + \Delta\theta\right) - T\cos\theta$$

Assume that the displacement in y is small and T is a constant along the stringthus θ and $\theta + \Delta\theta$ are smallthen $F_x \approx 0$ we can see this by expanding the trig functions

$$F_x \approx T[1 - \frac{\left(\theta + \Delta\theta\right)^2}{2} - 1 + \frac{\theta^2}{2} + \ldots]$$

or

$$F_x \approx T\theta\Delta\theta$$

which is very small.On the other hand

$$F_y \approx T[\theta + \Delta\theta - \theta + \ldots]$$

or

$$F_y \approx T\Delta\theta$$

which is not nearly as small. So we will consider the y component of motion, but approximate there is no x component

$$\begin{aligned} F_y &= T\sin(\theta + \Delta\theta) - T\sin\theta \\ &\approx T\tan(\theta + \Delta\theta) - T\tan\theta \\ &= T\left(\tfrac{\partial y(x+\Delta x)}{\partial x} - \tfrac{\partial y}{\partial x}\right) \\ &= T\tfrac{\partial^2 y}{\partial x^2}\Delta x \end{aligned}$$

Also we can write:

$$F_y = ma_y$$

$m = \mu\Delta x$ where μ is the mass density

$$a_y = \frac{\partial^2 y}{\partial t^2}$$

now have

$$T\frac{\partial^2 y}{\partial x^2}\Delta x = \mu\Delta x\frac{\partial^2 y}{\partial t^2}$$

$$\frac{\partial^2 y}{\partial x^2} = \frac{\mu}{T}\frac{\partial^2 y}{\partial t^2}$$

Note dimensions, get a velocity

$$\frac{T}{\mu} = \mathrm{v}^2$$

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{\mathrm{v}^2}\frac{\partial^2 y}{\partial t^2}$$

The second space derivative of a function is equal to the second time derivative of a function multiplied by a constant.

Normal Modes on a String

Before considering traveling waves, we are going to look at a special case solution to the wave equation. This is the case of stationary vibrations of a string.

For example here, lets consider the case where both ends of the string are fixed at $y = 0$. Now we vibrate the string. Every point along the string acts like a little driven oscillator. So lets assume that every point on string has a time dependence of the form $\cos \omega t$ and that the amplitude is a function of distance Assume

$$y(x,t) = f(x)\cos \omega t$$

then

$$\frac{\partial^2 y}{\partial t^2} = -\omega^2 f(x)\cos \omega t$$

$$\frac{\partial^2 y}{\partial x^2} = \frac{\partial^2 f}{\partial x^2}\cos \omega t$$

Substitute into wave equation

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{\mathrm{v}^2}\frac{\partial^2 y}{\partial t^2}$$

$$\frac{\partial^2 f}{\partial x^2}\cos \omega t = -\frac{\omega^2}{\mathrm{v}^2} f(x)\cos \omega t$$

Then every $f(x)$ that satisfies:

$$\frac{\partial^2 f}{\partial x^2} = -\frac{\omega^2}{\mathrm{v}^2} f$$

is a solution of the wave equation

A solution is (requiring $f(0) = 0$ since ends fixed)

$$f(x) = A \sin\left(\frac{\omega x}{\mathrm{v}}\right)$$

Another boundary condition is $f(L) = 0$ so get

$$A \sin\left(\frac{\omega L}{\mathrm{v}}\right) = 0$$

Thus

$$\frac{\omega L}{\mathrm{v}} = n\pi$$

$$\omega = \frac{n\pi \mathrm{v}}{L}$$

Be careful with the equations above: v is the letter vee and is for velocity. now we introduce the frequency ν which is the Greek letter nu.

recall $\nu = \omega/2\pi$ so

$$\nu_n = \frac{n\mathrm{v}}{2L} = \frac{n}{2L}\left(\frac{T}{\mu}\right)^{\frac{1}{2}}$$

This is a very important feature of wave phenomena. Things can be quantized. This is why a musical instrument will play specific notes. Note, that we must have an integral number of half sine waves

$$\lambda_n = \frac{2L}{n}$$

end up with

$$f_n(x) = A_n \sin\left(\frac{2\pi x}{\lambda_n}\right)$$

leading to

$$y_n(x,t) = A_n \sin\left(\frac{2\pi x}{\lambda_n}\right)\cos\omega_n t$$

where $\omega_n = \frac{n\pi}{L}\left(\frac{T}{\mu}\right)^{\frac{1}{2}} = \frac{n\pi}{L}\mathbf{v} = n\omega_1$ ω_1 is the fundamental frequency

Waves

Waves

The Wave Equation

In deriving the motion of a string under tension we came up with an equation:

$\frac{\partial^2 y}{\partial x^2} = \frac{1}{\mathrm{v}^2}\frac{\partial^2 y}{\partial t^2}$ which is known as the wave equation. We will show that this leads to waves below, but first, let us note the fact that solutions of this equation can be added to give additional solutions.

Waves Add

Say you have two waves governed by two equations Since they are traveling in the same medium, v is the same

$$\frac{\partial^2 f_1}{\partial x^2} = \frac{1}{\mathrm{v}^2}\frac{\partial^2 f_1}{\partial t^2}$$

$$\frac{\partial^2 f_2}{\partial x^2} = \frac{1}{\mathrm{v}^2}\frac{\partial^2 f_2}{\partial t^2}$$

add these

$$\frac{\partial^2 f_1}{\partial x^2} + \frac{\partial^2 f_2}{\partial x^2} = \frac{1}{\mathrm{v}^2}\frac{\partial^2 f_1}{\partial t^2} + \frac{1}{\mathrm{v}^2}\frac{\partial^2 f_2}{\partial t^2}$$

$$\frac{\partial^2}{\partial x^2}(f_1 + f_2) = \frac{1}{\mathrm{v}^2}\frac{\partial^2}{\partial t^2}(f_1 + f_2)$$

Thus $f_1 + f_2$ is a solution to the wave equation

Lets say we have two functions, $f_1(x - \mathrm{v}t)$ and $f_2(x + \mathrm{v}t)$. Each of these functions individually satisfy the wave equation. note that

$$y = f_1(x - \mathrm{v}t) + f_2(x + \mathrm{v}t)$$

will also satisfy the wave equation. In fact any number of functions of the form $f(x - \mathrm{v}t)$ or $f(x + \mathrm{v}t)$ can be added together and will satisfy the wave equation. This is a very profound property of waves. For example it will allow us to describe a very complex wave form, as the summation of simpler wave forms. The fact that waves add is a consequence of the fact that the wave equation

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{\mathrm{v}^2}\frac{\partial^2 f}{\partial t^2}$$

is linear, that is f and its derivatives only appear to first order. Thus any linear combination of solutions of the equation is itself a solution to the equation.

General Form

Any well behaved (ie. no discontinuities, differentiable) function of the form

$$y = f(x - \mathrm{v}t)$$

is a solution to the wave equation. Lets define

$$f'(a) = \frac{df}{da}$$

and

$$f''(a) = \frac{d^2 f}{da^2}.$$

Then using the chain rule

$$\frac{\partial y}{\partial x} = \frac{\partial f}{\partial (x-\mathrm{v}t)} \frac{\partial (x-\mathrm{v}t)}{\partial x}$$
$$= \frac{\partial f}{\partial (x-\mathrm{v}t)} = f'(x-\mathrm{v}t),$$

and

$$\frac{\partial^2 y}{\partial x^2} = f''(x-\mathrm{v}t).$$

Also

$$\frac{\partial y}{\partial t} = \frac{\partial f}{\partial (x-\mathrm{v}t)} \frac{\partial (x-\mathrm{v}t)}{\partial t}$$
$$= -\mathrm{v} \frac{\partial f}{\partial (x-\mathrm{v}t)}$$
$$= -\mathrm{v} f'(x-\mathrm{v}t)$$

$$\frac{\partial^2 y}{\partial t^2} = \mathrm{v}^2 f''(x-\mathrm{v}t).$$

We see that this satisfies the wave equation.

Lets take the example of a Gaussian pulse.

$$f(x-\mathrm{v}t) = Ae^{-(x-\mathrm{v}t)^2/2\sigma^2}$$

Then

$$\frac{\partial f}{\partial x} = \frac{-2(x-\mathrm{v}t)}{2\sigma^2} Ae^{-(x-\mathrm{v}t)^2/2\sigma^2}$$

and

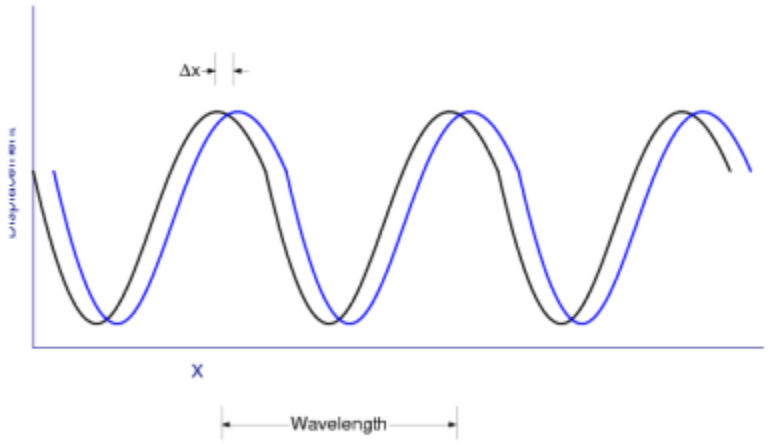
$$\frac{\partial f}{\partial t} = \frac{-2(x-\mathrm{v}t)(-\mathrm{v})}{2\sigma^2} Ae^{-(x-\mathrm{v}t)^2/2\sigma^2}$$

or

$$\frac{\partial^2 f(x-\mathrm{v}t)}{\partial t^2} = \mathrm{v}^2 \frac{\partial^2 f(x-\mathrm{v}t)}{\partial x^2}$$

That is it satisfies the wave equation.

The velocity of a Wave



To find the velocity of a wave, consider the wave:

$$y(x,t) = f(x-\mathrm{v}t)$$

Then can see that if you increase time and x by Δt and Δx for a point on the traveling wave of constant amplitude

$$f(x-\mathrm{v}t) = f\left((x+\Delta x) - \mathrm{v}(t+\Delta t)\right).$$

Which is true if

$$\Delta x - \mathrm{v}\Delta t = 0$$

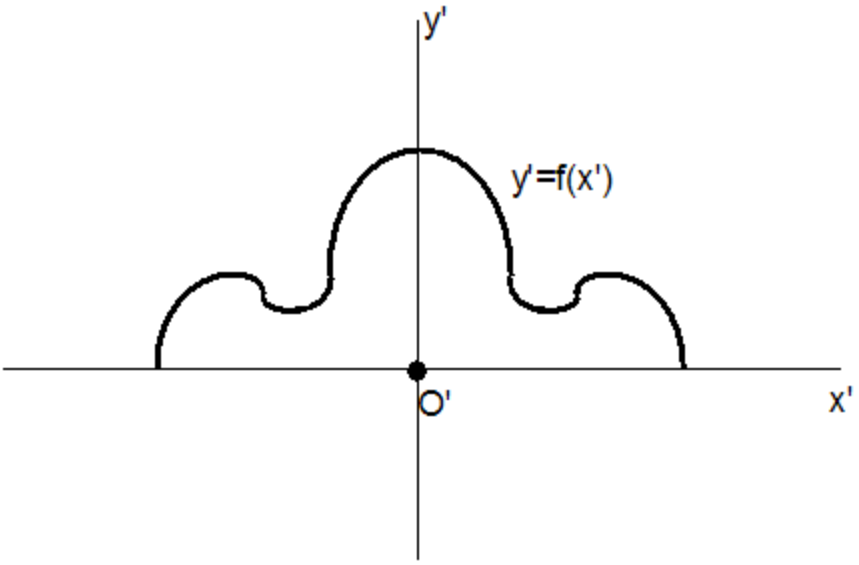
or

$$\mathrm{v} = \frac{\Delta x}{\Delta t}$$

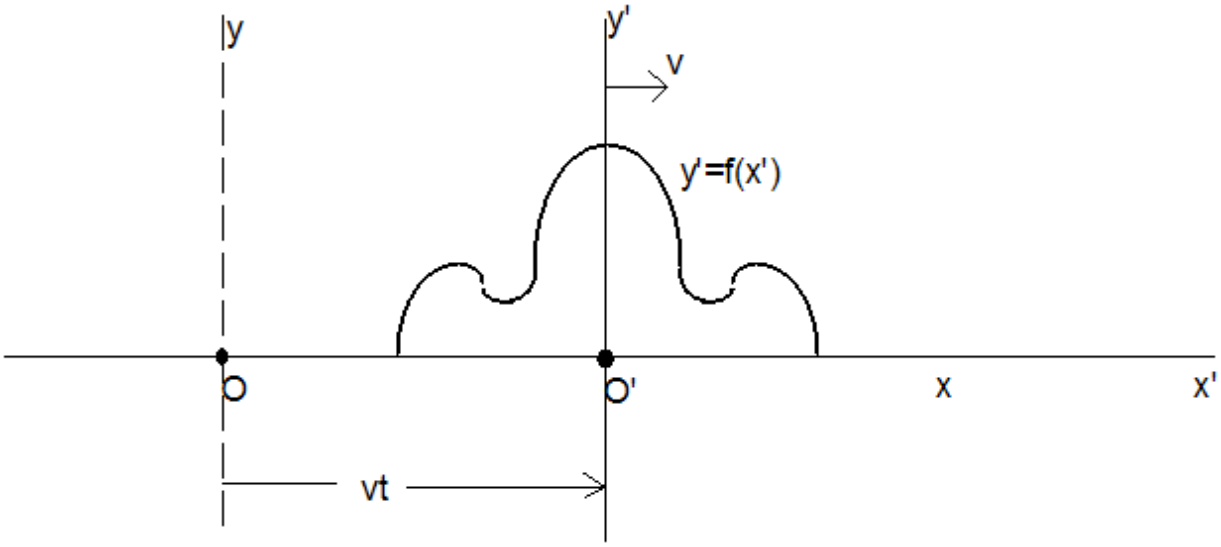
Thus $f(x - \mathrm{v}t)$ describes a wave that is moving in the positive x direction. Likewise $f(x + \mathrm{v}t)$ describes a wave moving in the negative x direction.

Note: Lots of students get this backwards so watch out!

Another way to picture this is to consider a one dimensional wave pulse of arbitrary shape, described by $y' = f(x')$, fixed to a coordinate system $O'(x', y')$



Now let the O' system, together with the pulse, move to the right along the x-axis at uniform speed v relative to a fixed coordinate system $O(x, y)$.



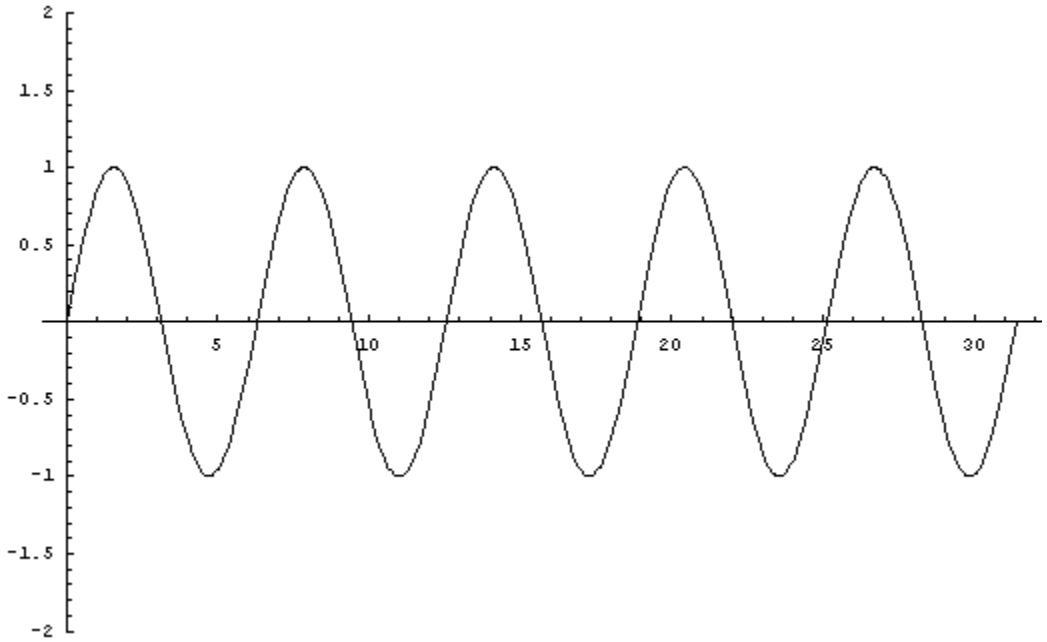
As it moves, the pulse is assumed to maintain its shape. Any point P on the pulse can be described by either of two coordinates x or x', where $x' = x - \mathrm{v}t$. The y coordinate is identical in either system. In the stationary coordinate system's frame of reference, the moving pulse has the mathematical form

$$y = y' = f(x') = f(x - \mathrm{v}t)$$

If the pulse moves to the left, the sign of v must be reversed, so that we may write

$$y = f(x \pm \mathrm{v}t)$$

as the general form of a traveling wave. Notice that we have assumed $x = x'$ at $t = 0$.



Note: Waves carry momentum, energy (possibly angular momentum) but not matter

Wavelength, Wavenumber etc.

We will often use a sinusoidal form for the wave. However we can't use

$$y = A \sin{(x - \mathrm{v}t)}$$

since the part in brackets has dimensions of length. Instead we use

$$y = A \sin{\frac{2\pi}{\lambda}(x - \mathrm{v}t)}.$$

Notice that

$$y(x = 0, t) = y(x = \lambda, t)$$

which gives us the definition of the wavelength λ.

Also note that the frequency is

$$\nu = \frac{\mathrm{v}}{\lambda}.$$

The angular frequency is defined to be

$$\omega \equiv 2\pi\nu = \frac{2\pi\mathrm{v}}{\lambda}.$$

Finally the wave number is

$$k \equiv \frac{2\pi}{\lambda}.$$

So we could have written our wave as

$$y = A\sin\left(kx - \omega t\right)$$

Note that some books say $k = \frac{1}{\lambda}$

Normal Modes on a String as an Example of Wave Addition

Lets go back to our solution for normal modes on a string:

$$y_n(x,t) = A_n \sin\left(\frac{2\pi x}{\lambda_n}\right)\cos\omega_n t$$

$$y_n(x,t) = A_n \sin\left(\frac{2\pi x}{\lambda_n}\right)\cos\left(\frac{2\pi}{\lambda_n}\mathrm{v}t\right).$$

Now lets do the following: make use of
$\sin\left(\theta+\varphi\right)+\sin\left(\theta-\varphi\right) = 2\sin\theta\cos\varphi$ Also lets just take the first normal mode and drop the n's Finally, define $A \equiv A_1/2$ Then

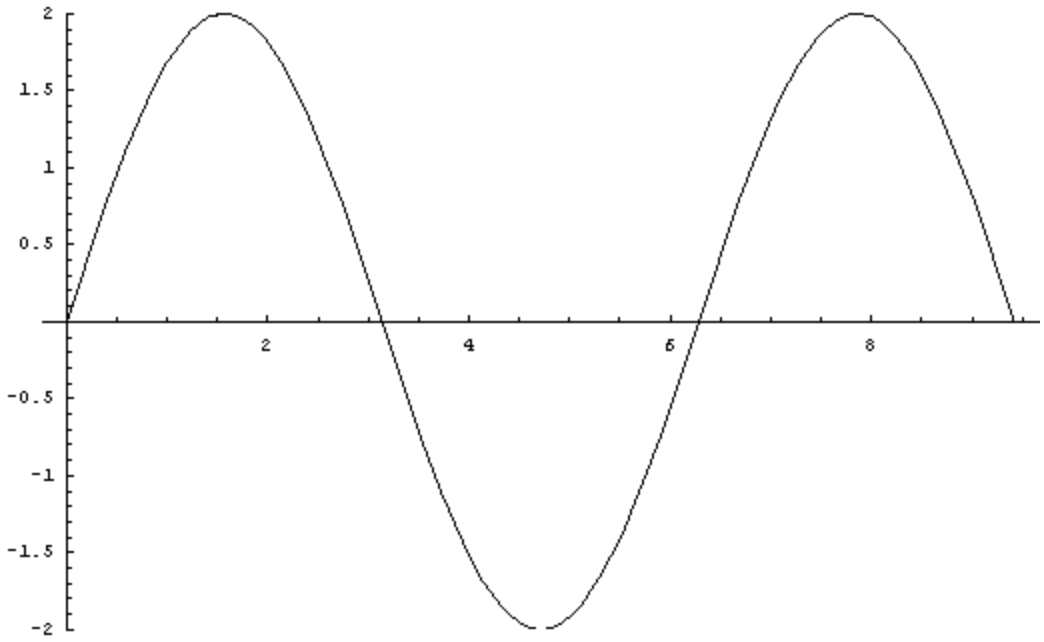
$$y(x,t) = 2A\sin\left(\frac{2\pi x}{\lambda}\right)\cos\left(\frac{2\pi}{\lambda}\mathrm{v}t\right)$$

becomes

$$y(x,t) = A \sin\left[\frac{2\pi}{\lambda}(x - \mathrm{v}t)\right] + A \sin\left[\frac{2\pi}{\lambda}(x + \mathrm{v}t)\right]$$

These are two waves of equal amplitude and speed traveling in opposite directions. We can plot what happens when we do this. The following animation was made with Mathematica using the command

```
Do[Plot[(Sin[x - 0.5 t] + Sin[x + 0.5 t]), {x, 0, 3 Pi}, PlotRange → {-2, 2}], {t, 0, 4 Pi, 0.2}]
```



Module 10.xhtml

Superposition

Suppose we have two waves, with the same amplitude but different wavelengths and velocities and we add them

$$y_1 = A \sin\left[\frac{2\pi}{\lambda_1}\left(x - \mathrm{v}_1 t\right)\right]$$

$$y_2 = A \sin\left[\frac{2\pi}{\lambda_2}\left(x - \mathrm{v}_2 t\right)\right].$$

Then

$$y_1 + y_2 = A\left(\sin\left[\frac{2\pi}{\lambda_1}\left(x - \mathrm{v}_1 t\right)\right] + \sin\left[\frac{2\pi}{\lambda_2}\left(x - \mathrm{v}_2 t\right)\right]\right).$$

Lets rewrite using wave number and angular frequency

$$y_1 + y_2 = y = A\left(\sin\left[(k_1 x - \omega_1 t)\right] + \sin\left[(k_2 x - \omega_2 t)\right]\right).$$

Now we will use $\sin(\theta + \varphi) + \sin(\theta - \varphi) = 2\sin\theta\cos\varphi$ and set

$$\theta + \varphi = k_1 x - \omega_1 t$$

$$\theta - \varphi = k_2 x - \omega_2 t.$$

We can rearrange to get

$$2\theta = (k_1 + k_2)x - (\omega_1 + \omega_2)t$$

$$2\varphi = (k_1 - k_2)x - (\omega_1 - \omega_2)t.$$

By substituting we can then see that

$$y = 2A\left(\cos\left[\frac{k_1 - k_2}{2}x - \frac{\omega_1 - \omega_2}{2}t\right] \times \sin\left[\frac{k_1 + k_2}{2}x - \frac{\omega_1 + \omega_2}{2}t\right]\right).$$

Now set

$$\Delta k = k_1 - k_2$$

$$\Delta \omega = \omega_1 - \omega_2$$

$$k = \frac{k_1 + k_2}{2}$$

$$\omega = \frac{\omega_1 + \omega_2}{2}$$

and we can rewrite the wave as

$$y = 2A\cos\left(x\frac{\Delta k}{2} - t\frac{\Delta \omega}{2}\right)\sin\left(kx - \omega t\right).$$

The above equation shows beats. For example you can set $t = 0$ and see that you get

$$y = 2A\cos\left(x\frac{\Delta k}{2}\right)\sin\left(kx\right).$$

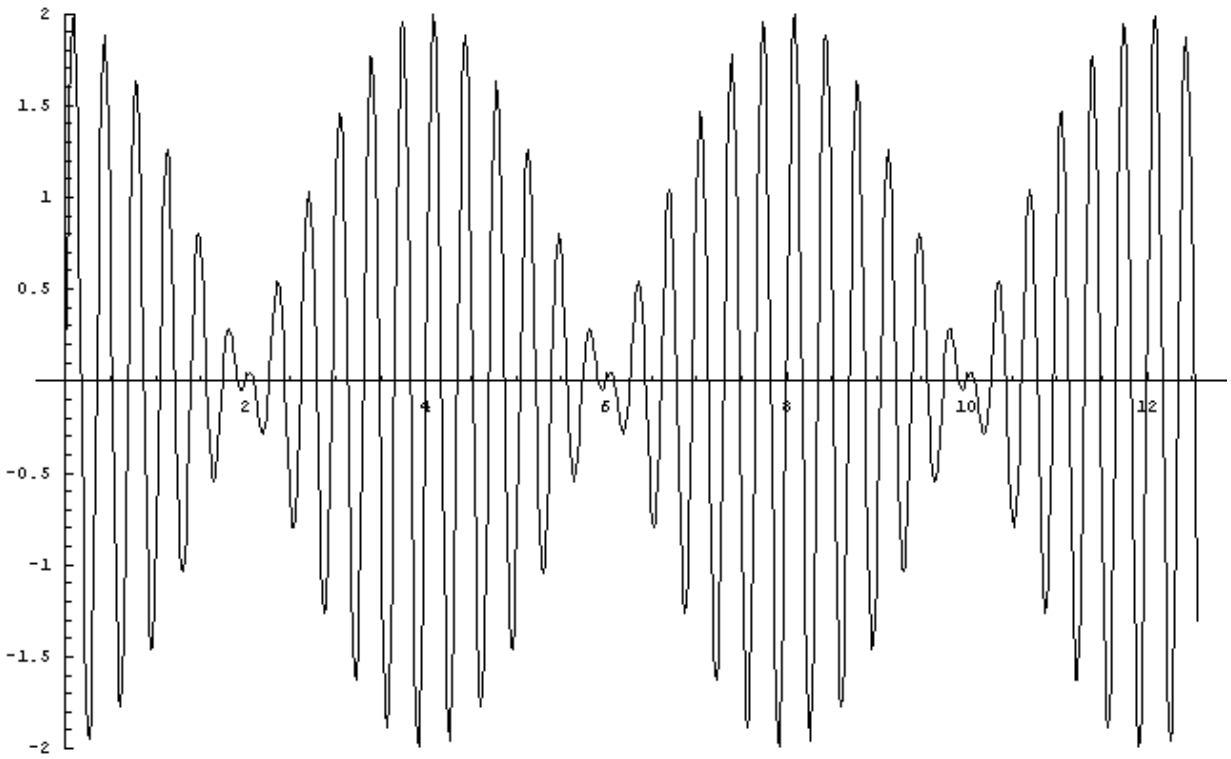
Likewise you could pick $x = 0$ and get the same figure, but now the horizontal axis is time

$$y = 2A\cos\left(-t\frac{\Delta \omega}{2}\right)\sin\left(-\omega t\right)$$

or

$$y = 2A\cos\left(t\frac{\Delta \omega}{2}\right)\sin\left(-\omega t\right).$$

You get a traveling wave that has an oscillating amplitude.



Adding two waves of similar frequency together, gives rise to beats.

Phase and Group Velocities

When we look at

$$y = 2A\cos\left(x\frac{\Delta k}{2} - t\frac{\Delta\omega}{2}\right)\sin\left(kx - \omega t\right)$$

we see that there are two velocities. One, referred to as the phase velocity, is the speed of the individual wave crests:

$$\mathbf{v}_p = \frac{\omega}{k} = \nu\lambda.$$

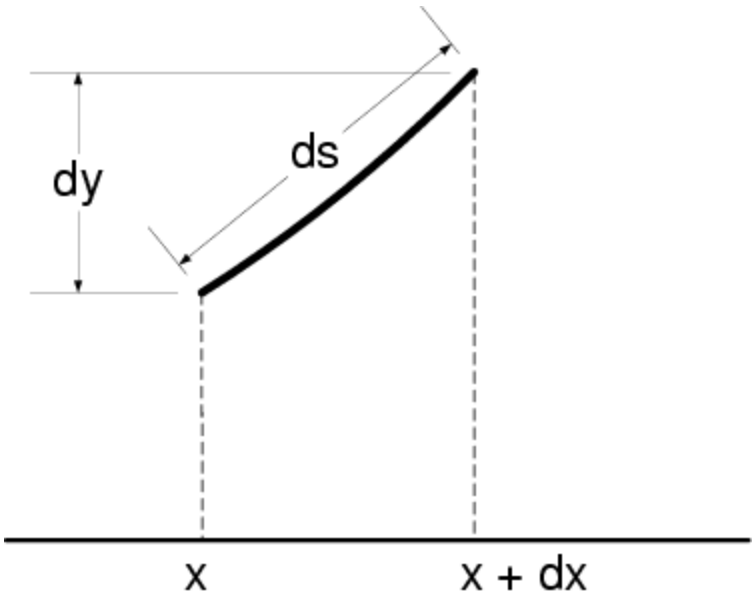
The group velocity is the velocity of the envelope

$$\mathbf{v}_g = \frac{\Delta\omega}{\Delta k} \rightarrow \frac{d\omega}{dk}$$

Energy and momentum normally move with the group velocity.

Energy in a mechanical wave

Energy Transport



Lets calculate the energy in a wave of a string:

Consider a fragment of string so small it can be considered straight, as is shown in the figure

The the kinetic energy is $K = \frac{1}{2}m\mathrm{v}^2$ for the string fragment $m = \mu \mathbb{d}x$

Note: Why is this $\mathbb{d}x$ and not $\mathbb{d}s$? Lets consider that the string is not perturbed, then it is horizontal and has mass as given. When the string is perturbed it stretches a little bit - but the mass does not increase.

So we have

$$K = \frac{1}{2}\mu \mathbb{d}x \left(\frac{\partial y}{\partial t} \right)^2$$

and using this we can define the energy per unit length, ie. the kinetic energy density:

$$\frac{\mathbb{d}K}{\mathbb{d}x} = \frac{1}{2}\mu \left(\frac{\partial y}{\partial t} \right)^2$$

When the string segment is stretched from the length $\mathbb{d}x$ to the length $\mathbb{d}s$ an amount of work $= T(\mathbb{d}s - \mathbb{d}x)$ is done. This is equal to the potential energy stored in the stretched string segment. So the potential energy in this case is:

$$U = T(\mathbb{d}s - \mathbb{d}x)$$

Now

$$\begin{aligned} \mathbb{d}s &= (\mathbb{d}x^2 + \mathbb{d}y^2)^{1/2} \\ &= \mathbb{d}x \left[1 + \left(\tfrac{\partial y}{\partial x} \right)^2 \right]^{1/2} \end{aligned}$$

Recall the binomial expansion

$$(1 + A)^n = 1 + nA + \frac{n(n-1)A^2}{2!} + \frac{n(n-1)(n-2)A^3}{3!} + \dots$$

so

$$\mathbb{d}s \approx \mathbb{d}x + \frac{1}{2} \left(\frac{\partial y}{\partial x} \right)^2 \mathbb{d}x$$

$$U = T(\mathbb{d}s - \mathbb{d}x) \approx \frac{1}{2} T \left(\frac{\partial y}{\partial x} \right)^2 \mathbb{d}x$$

or the potential energy density

$$\frac{\mathrm{d}U}{\mathrm{d}x} = \frac{1}{2}T\left(\frac{\partial y}{\partial x}\right)^2$$

To get the kinetic energy in a wavelength, lets start with

$$y = A\sin\left(\frac{2\pi x}{\lambda} - \omega t\right)$$

$$\frac{\partial y}{\partial t} = -\omega A\cos\left(\frac{2\pi x}{\lambda} - \omega t\right)$$

Lets evaluate it at time 0.

$$\left.\frac{\partial y}{\partial t}\right|_{t=0} = -\omega A\cos\left(\frac{2\pi x}{\lambda}\right)$$

so

$$\frac{\mathrm{d}K}{\mathrm{d}x} = \frac{1}{2}\mu\omega^2 A^2\cos^2\left(\frac{2\pi x}{\lambda}\right)$$

now integrate

$$K = \int_0^\lambda \frac{\mathrm{d}K}{\mathrm{d}x}\mathrm{d}x$$
$$= \tfrac{1}{2}\mu\omega^2 A^2\int_0^\lambda \cos^2\left(\tfrac{2\pi x}{\lambda}\right)\mathrm{d}x$$

In order to do this integral we use the following trig identity:

$$\cos^2 A = \frac{\cos 2A + 1}{2}$$

so we get

$$K = \tfrac{1}{2}\mu\omega^2 A^2\left[\tfrac{x}{2} + \tfrac{\lambda}{8\pi}\sin\tfrac{4\pi x}{\lambda}\right]\Big|_{x=0}^{\lambda}$$
$$= \tfrac{1}{4}\mu\lambda\omega^2 A^2$$

In similar fashion the potential energy can be found to be

$$U = \frac{1}{4}\mu\lambda\omega^2 A^2.$$

Deriving this will be assigned as a homework problem

So

$$E = K + U = \frac{1}{2}\mu\lambda\omega^2 A^2$$

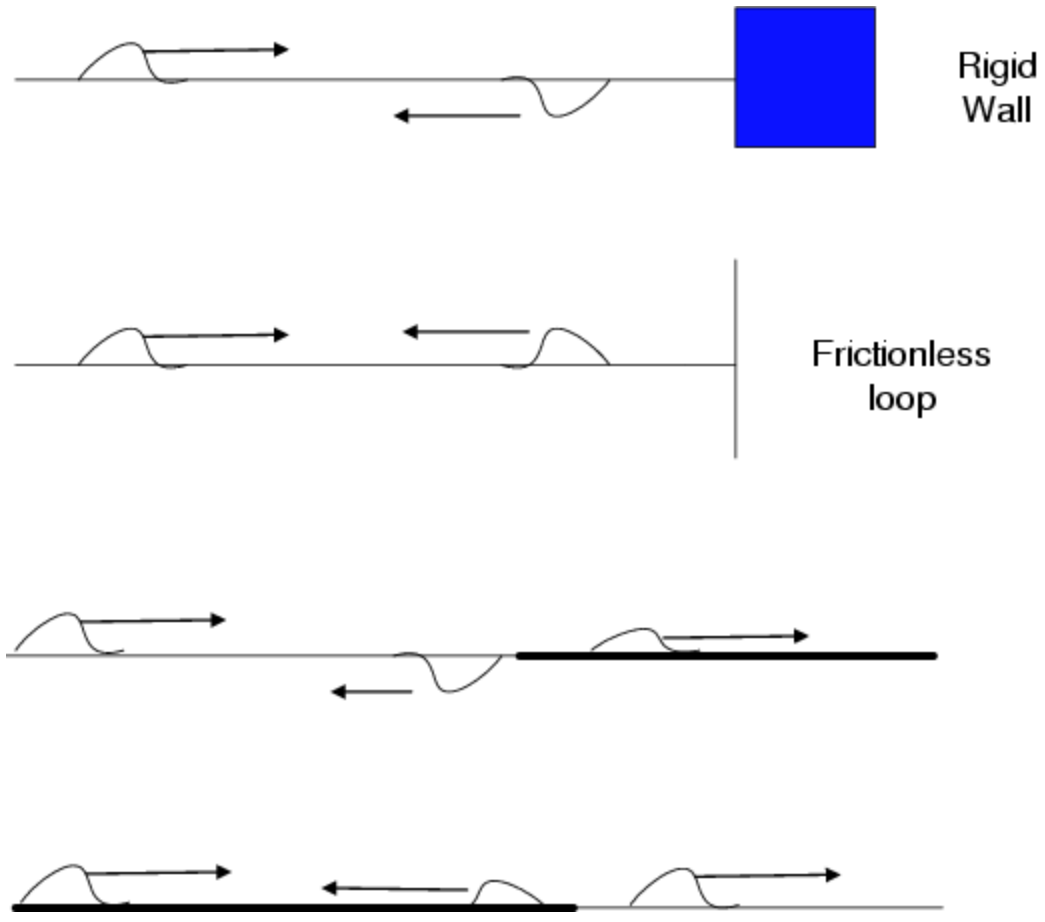
Power

$$\begin{aligned} P = \frac{\Delta E}{\Delta t} &= \frac{\frac{1}{2}\mu\lambda\omega^2 A^2}{\tau} \\ &= \tfrac{1}{2}\mu\omega^2 A^2 \mathbf{v} \end{aligned}$$

Where I have used $\tau = 1/\nu$ and $\lambda\nu =$v thus $\tau = \lambda/\mathbf{v}$

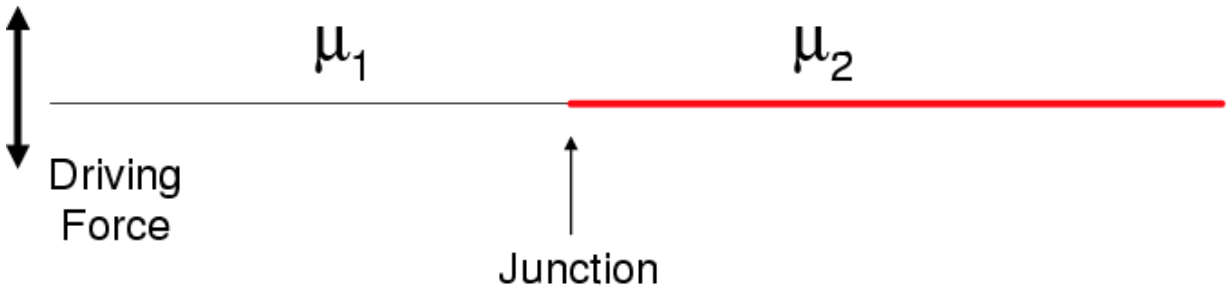
Module 13.xhtml

Reflection and Transmission



Four possible cases for the reflection and transmission of a wave on a string.

The first figure shows the 4 possible cases for reflection and transmission at an interface. Lets solve the problem, which is shown in the next figure



since $\mu_1 \neq \mu_2$ it must be that $\mathrm{v}_1 \neq \mathrm{v}_2$.

Note: Note that we are assuming that Young's Modulus is constant across the boundary.

So we get

$$y_{inc} = A\cos\left(k_1 x - \omega t\right)$$

$$y_{ref} = B\cos\left(k_1 x + \omega t\right)$$

$$y_{trans} = C\cos\left(k_2 x - \omega t\right)$$

(note the reflected wave goes the other direction).

On the left side of the junction we have

$$\begin{aligned} y_l &= y_{inc} + y_{ref} \\ &= A\cos\left(k_1 x - \omega t\right) + B\cos\left(k_1 x + \omega t\right) \end{aligned}$$

and on the right side of the junction we have

$$y_r = y_{trans} = C\cos\left(k_2 x - \omega t\right).$$

At the boundary $x = 0$ the wave must continuous, that is there are no kinks in it. Thus we must have

$$y_l(0,t) = y_r(0,t)$$

$$\left.\frac{\partial y_l(x,t)}{\partial x}\right|_{x=0} = \left.\frac{\partial y_r(x,t)}{\partial x}\right|_{x=0}$$

So from the first equation

$$A\cos(\omega t) + B\cos(\omega t) = C\cos(\omega t)$$

$$A + B = C$$

$$\left.\frac{\partial y_l(x,t)}{\partial x}\right|_{x=0} = \left.\frac{\partial y_r(x,t)}{\partial x}\right|_{x=0}$$

$$-Ak_1\sin(-\omega t) - k_1 B\sin(\omega t) = -k_2 C\sin(-\omega t)$$

$$(A-B)k_1\sin\omega t = Ck_2\sin\omega t$$

$$A - B = \frac{k_2}{k_1}C$$

now solve for B and C

$$A + B = C$$

$$A - B = \frac{k_2}{k_1}C$$

$$2A = \left(1 + \frac{k_2}{k_1}\right)C$$

Thus we can define the transmission coefficient

$$t_r \equiv C/A = \frac{2k_1}{k_1 + k_2}$$

and the refection coefficient

$$r \equiv B/A = \frac{C}{A} - 1 = \frac{k_1 - k_2}{k_1 + k_2}$$

note how the amplitudes can change at the boundary

If $\mu_2 < \mu_1$then we must have $k_2 < k_1$ since

$$\mathrm{v} = \omega/k = \sqrt{T/\mu}$$

and ω and$T$ must be fixed. We see that $k \propto \sqrt{\mu}$ In this case we see that the amplitude of the wave gets bigger when it moves into a less dense medium. We have probably all experienced this in real life. As waves come ashore they become bigger. This is because shallower water is effectively less dense. A tsumami in open ocean may have an imperceptable amplitude but when it comes ashore it can be tremendous. This seems almost counter intuitive, but in any closed system the energy and power are conserved but there is no rule saying amplitude has to be conserved.

Lets look at the reflected and transmitted power. Recall Power:

$$P = \frac{1}{2}\mu\omega^2 A^2 \mathrm{v}$$

For the incident and reflected waves μ and ν are the same so the reflected power coefficient (reflected power / incident power)

$$P_R = (B/A)^2 = \left(\frac{k_1 - k_2}{k_1 + k_2}\right)^2$$

To do transmitted power lets first rewrite the power equation. Recall

$$\begin{aligned} \mathrm{v} &= \nu\lambda \\ &= \tfrac{2\pi\nu}{k} \\ &= \tfrac{\omega}{k} \end{aligned}$$

Also

$$\mathrm{v} = \sqrt{\frac{T}{\mu}}$$

so

$$\begin{aligned}\mu &= \tfrac{T}{\mathrm{v}^2} \\ &= \tfrac{T}{\left(\frac{\omega}{k}\right)^2} \\ &= \tfrac{Tk^2}{\omega^2}\end{aligned}$$

so now

$$P = \frac{1}{2}\mu\omega^2 A^2 \mathrm{v}$$

becomes

$$P = \frac{1}{2}\frac{Tk^2}{\omega^2}\omega^2 A^2 \frac{\omega}{k}$$

or

$$P = \frac{1}{2}Tk\omega A^2.$$

Note: Watch out, in the above lines A was used to denote amplitude in general and in the following line it specifically refers to the incoming wave.

The transmission power coefficient is thus:

$$P_T = \frac{\frac{1}{2}Tk_2\omega C^2}{\frac{1}{2}Tk_1\omega A^2}$$

Note that ω and T are the same for both waves

$$P_T = \frac{k_2 C^2}{k_1 A^2}$$

earlier we showed

$$C/A = \frac{2k_1}{k_1 + k_2}$$

so

$$P_T = \left(\frac{2k_1}{k_1 + k_2} \right)^2 \frac{k_2}{k_1}$$

$$P_T = \frac{4k_1 k_2}{(k_1 + k_2)^2}$$

Note that $P_R + P_T = 1$ which means that energy is conserved.

Now lets look at the 4 specific cases we have:

Rigid wall

$\mu \to \infty$ so $k_2 \to \infty$

$$\begin{aligned} r &= \frac{k_1 - k_2}{k_1 + k_2} \\ &= \frac{\frac{k_1}{k_2} - 1}{\frac{k_1}{k_2} + 1} \\ r &\to -1 \end{aligned}$$

Also $P_R \to +1$ $P_T \to 0$ So wave is reflected and inverted, but has same power

Free end

$\mu \to 0$ so $k_2 \to 0$

$$
\begin{aligned}
r &= \frac{k_1 - k_2}{k_1 + k_2} \\
&= \frac{k_1}{k_1} \\
r &\to +1
\end{aligned}
$$

Also $P_R \to +1$ $P_T \to 0$ So wave is reflected and has same power

Moving to higher density

$\mu_2 > \mu_1 \quad k_2 > k_1$ so $r < 0$ $t_r > 0$

Moving to lower density

$\mu_2 < \mu_1 \quad k_2 < k_1$ so $r > 0$ $t_r > 1$ Note the transmitted wave's amplitude is larger than the original.

Fourier Series

Fourier Analysis

Fourier Series

Lets go back to the case of a string fixed at 0 and L, its n^{th} harmonic is

$$y_n(x,t) = A_n \sin\left(\frac{n\pi x}{L}\right)\cos\left(\omega_n t - \delta_n\right)$$

In fact all the modes could be permitted, and so any possible motion of the string can be completely specified by:

$$y(x,t) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{L}\right)\cos\left(\omega_n t - \delta_n\right).$$

This has been rigorously shown by mathematicians but the complete proof is beyond our scope in this course. Lets accept the mathematicians word on this. We could take a snapshot of this function at a time $t = t_0$. Then we could write

$$y(x) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right)$$

where

$$B_n = A_n \cos\left(\omega_n t_0 - \delta_n\right).$$

Likewise we could look at one point at space and look at the oscillations as a function of time. In that case we would get.

$$y(t) = \sum_{n=1}^{\infty} C_n \cos\left(\omega_n t - \delta_n\right)$$

Lets work with the time snapshot,

$$y(x) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right)$$

We need to figure out what the B_n factors are and this is what Fourier figured out. We can multiply both sides by the $\sin$ of a particular harmonic

$$y(x) sin\left(\frac{n_i\pi x}{L}\right) = \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) sin\left(\frac{n_i\pi x}{L}\right)$$

and now we can integrate both sides Recall

$$\cos(\theta - \varphi) = \cos\theta\cos\varphi + \sin\theta\sin\varphi$$

$$\cos(\theta + \varphi) = \cos\theta\cos\varphi - \sin\theta\sin\varphi$$

So

$$\sin\theta\sin\varphi = \frac{1}{2}\left[cos(\theta - \varphi) - \cos(\theta + \varphi)\right]$$

Thus This is equal to zero at the limits $0, L$ except for the particular case when $n = n_i$. In that case

$$\int \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{n_i\pi x}{L}\right) dx = \int \sin^2\left(\frac{n\pi x}{L}\right) dx$$

So you get After all that we should see that for each term in the sum is zero, except the case where $n_i = n$. Thus we can simplify the equation:

$$\int_0^L y(x) \sin\left(\frac{n\pi x}{L}\right) dx = \frac{L}{2} B_n.$$

or

$$B_n = \frac{2}{L}\int_0^L y(x) \sin\left(\frac{n\pi x}{L}\right) dx$$

The above is a very specific form of the Fourier Series for a function spanning an interval from 0 to L and passing through zero at $x = 0$.

More General Case

One could write a more general case for the Fourier Series which applies to an interval spanning $-L$ to L and not constrained to pass through zero. In that case one can write

$$y(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right]$$

where

$$A_n = \frac{1}{L} \int_{-L}^{L} y(x) \cos\left(\frac{n\pi x}{L}\right) dx \qquad n = 0, 1, 2, 3, \ldots$$

and

$$B_n = \frac{1}{L} \int_{-L}^{L} y(x) \sin\left(\frac{n\pi x}{L}\right) dx \qquad n = 1, 2, 3, \ldots$$

You can then look at the symmetry of the problem and see if just $\sin$ or $\cos$ can be used. For example if $y(-x) = y(x)$ then use cosines. If $y(-x) = -y(x)$ use the sines.

Fourier Integral Theorem

In fact Fourier's theorem can be taken to a next step. This is Fourier's integral theorem. That is any function (even if it is not periodic) can be represented by

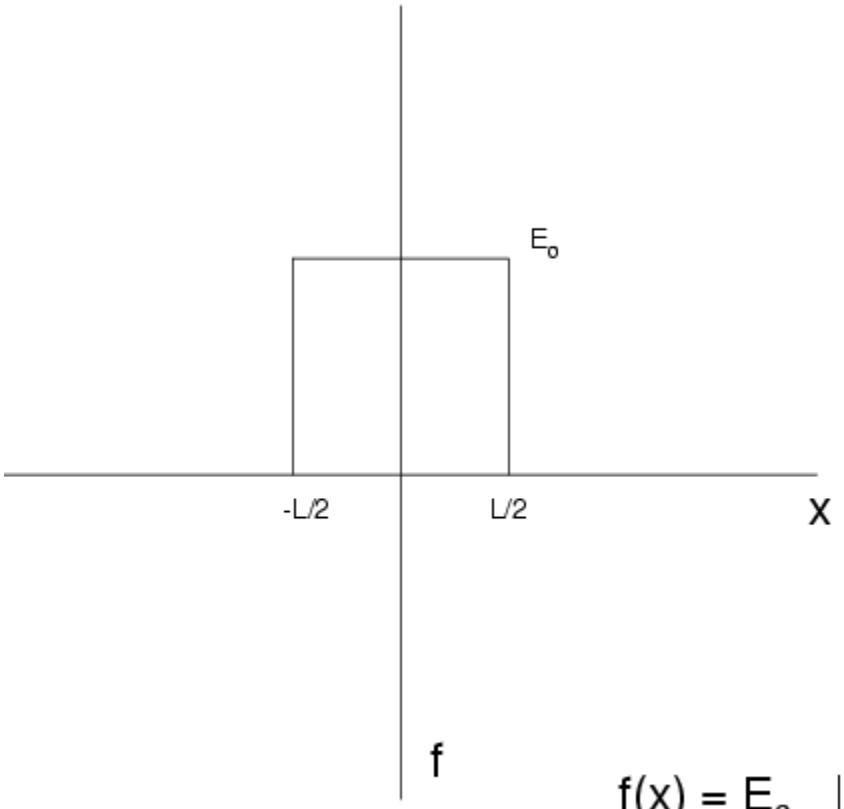
$$f(x) = \frac{1}{\pi} \int_0^{\infty} [A(k) \cos(kx) + B(k) sin(kx)] dk$$

where

$$A(k) = \int_{-\infty}^{\infty} f(x) \cos(kx) \mathbb{d}x$$

$$B(k) = \int_{-\infty}^{\infty} f(x) \sin(kx) \mathbb{d}x$$

A and B are called the Fourier transforms of $f(x)$ Lets look at an example.



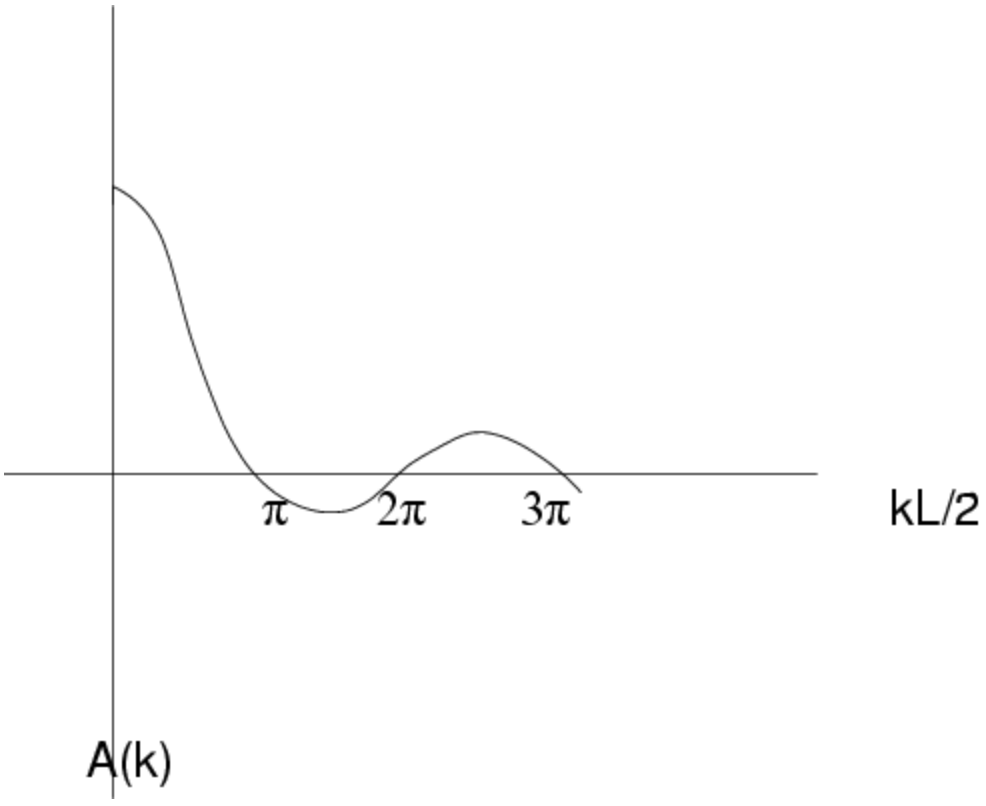
$$f(x) = E_o \quad |x| < L/2$$
$$f(x) = 0 \quad |x| > L/2$$

right away you can set

$$B(x) = 0$$

from symmetry arguments

$$\begin{aligned} A(k) &= \int_{-\infty}^{\infty} f(x) \cos(kx) \mathrm{d}x \\ &= \int_{-L/2}^{L/2} E_0 \cos(kx) \mathrm{d}x \\ &= \frac{E_o}{k} \sin(kx) \Big|_{-L/2}^{L/2} \\ &= \frac{E_o}{k} \left[\sin\left(\frac{kL}{2}\right) - \sin\left(\frac{-kL}{2}\right) \right] \\ &= \frac{2E_o}{k} \sin\left(\frac{kL}{2}\right) \\ &= E_0 L \frac{\sin\left(\frac{kL}{2}\right)}{\frac{kL}{2}} \end{aligned}$$



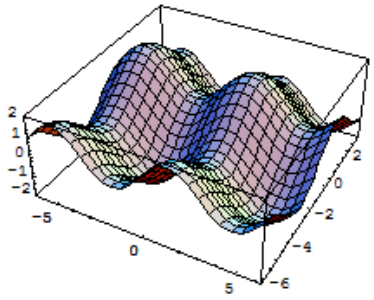
Closing word

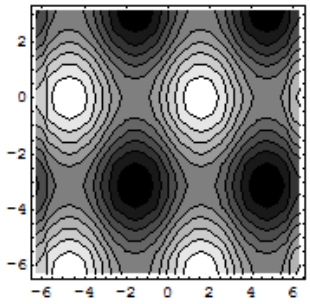
Up until now in the course we have been dealing with very simple waves. It turns out that any complicated wave that can possibly exist can be constructed from simple harmonic waves. So while it may seem that an harmonic wave is an over simplification, it can be used in even the most complex cases.

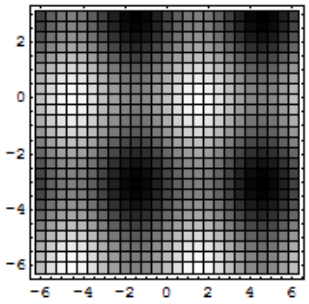
Scalar.xhtml

Scalar Fields

One of the more difficult concepts we encounter in physics is the notion of a field. However it is an extremely useful concept. A scalar field is a map over some space of scalar values. That is it is a map of values with no direction. A simple example of a scalar field is a map of the temperature distribution in a room. In this course the most important example is the electromagnetic potential field. Below are a few examples of graphical representations of one particular scalar field.







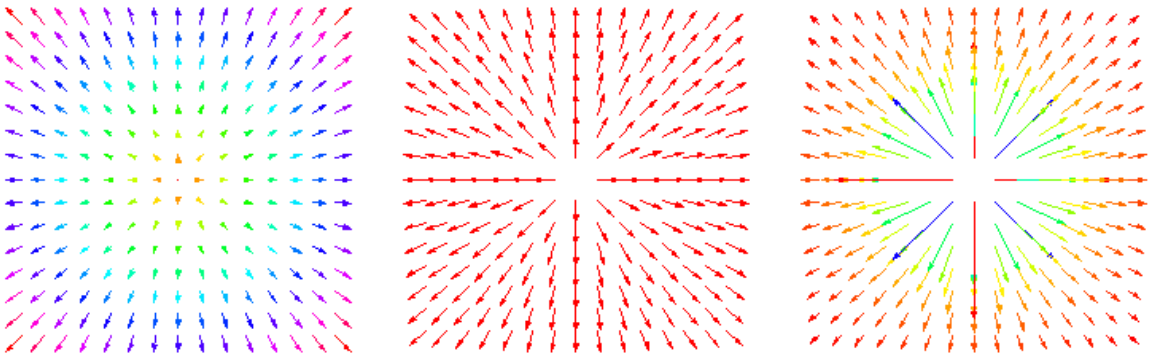
Vector Fields

Vector Fields

A vector field can be considered a map of vectors over some space. . For example if one were to show wind vectors on a weather map; that would be a vector field. The electric field surrounding a charge is a vector field (were is the potential around the charge is a scalar field).

The flux of a vector field through a closed surface is the average outward normal component of the vector times the area of the surfaceflux = (average normal component)·(surface area)

Three vector fields are shown below. Which represents the electric field eminating from a positive point charge in the middle? (Note that vectors of similar magnitude are colored similarly in these plots)



Vector multiplication.xhtml

It is presumed that you are familiar with vector multiplication.

Here are some questions to ask yourself as a refresher:

$\vec{A} \cdot \vec{B}$. Is it a vector or a scalar?

Answer: scalar

$$\vec{A} \cdot \vec{B} = A_x B_x + A_y B_y + A_z B_z$$

$\vec{A} \times \vec{B}$. Is it a vector or a scalar?

Answer: Vector

$$\begin{aligned}(\vec{A} \times \vec{B})_x &= A_y B_z - A_z B_y \\ (\vec{A} \times \vec{B})_y &= A_z B_x - A_x B_z \\ (\vec{A} \times \vec{B})_z &= A_x B_y - A_y B_x\end{aligned}$$

What is $\vec{A} \times \vec{A}$

Answer: 0

What is $\vec{A} \cdot (\vec{A} \times \vec{B})$

Answer: 0

Another useful thing to remember

$$\vec{A} \cdot (\vec{B} \times \vec{C}) = (\vec{A} \times \vec{B}) \cdot \vec{C}$$

$$= (\vec{C} \times \vec{A}) \cdot \vec{B}$$

This is the scalar triple product which is the volume of a parallelopiped whose edges are given by $\vec{A},\ \vec{B},\ \vec{C}$.

Gradient.xhtml

Divergence, Gradient and Curl

Assume we have measured the temperature in a room along an axis x. If we wanted to find the temperature change as we move to postion $(x + \Delta x)$ then from the fundamental definition of a derivative we know that is:

$$\Delta T = \frac{\mathbb{d}T}{\mathbb{d}x}\Delta x$$

We can easily extend this concept to 3 dimensions At position (x, y, z) there is a temperature $T(x, y, z)$. Suppose we then want to find the temperature at $\vec{R} + \Delta\vec{R} = (x + \Delta x, y + \Delta y, z + \Delta z)$. Then we can use:

$$\Delta T = \frac{\partial T}{\partial x}\Delta x + \frac{\partial T}{\partial y}\Delta y + \frac{\partial T}{\partial z}\Delta z$$

We could define a vector

$$\left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}, \frac{\partial T}{\partial z}\right)$$

and then say

$$\Delta T = \left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}, \frac{\partial T}{\partial z}\right) \cdot \Delta\vec{R}$$

so let's define an operator

$$\vec{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$$

Then we can write

$$\Delta T = \vec{\nabla}T \cdot \Delta\vec{R}$$

$\vec{\nabla}$ is a vector operator that can be used in other situations involving scalars and vectors. It is often named "del" or "nabla". Operating on a scalar field with this operator is called taking the "gradient" of the field.

We could also operate on a vector field with del. There are two different ways to do this, by taking the dot and the cross products. To operate on a vector field by taking its dot product with del is referred to as taking the divergence. ie.

$$f = \vec{\nabla} \cdot \vec{h}$$

where $\vec{h}$ is some vector field and f is the resulting scalar field.

Similarly one could take the cross product:

$$\vec{g} = \vec{\nabla} \times \vec{h}$$

where $\vec{g}$ is the resulting vector field.

$$\begin{aligned} g_x &= \left(\vec{\nabla} \times \vec{h}\right)_x = \frac{\partial h_z}{\partial y} - \frac{\partial h_y}{\partial z} \\ g_y &= \left(\vec{\nabla} \times \vec{h}\right)_y = \frac{\partial h_x}{\partial z} - \frac{\partial h_z}{\partial x} \\ g_z &= \left(\vec{\nabla} \times \vec{h}\right)_z = \frac{\partial h_y}{\partial x} - \frac{\partial h_x}{\partial y} \end{aligned}$$

This is referred to as taking the curl of a field.

These operations, Gradient, Divergence and Curl are of fundamental importance. They have been presented above as operations using some newly defined operator but they in fact have deep physical significance. When using these operators to express Maxwell's equations in differential form, the meaning of these operations will hopefully become more clear. Gradient is the easiest to understand, it can be thought of as a three dimensional slope.

Having defined these operations we can go on to second derivative type things

$$\vec{\nabla} \cdot (\vec{\nabla} T) = \nabla^2 T = a\ scalar\ field$$

Note that ∇^2 occurs so often that is has its own name, Laplacian

$$\vec{\nabla} \times (\vec{\nabla} T) = 0$$

$$\vec{\nabla} (\vec{\nabla} \cdot \vec{h}) = a\ vector\ field$$

$$\vec{\nabla} \cdot (\vec{\nabla} \times \vec{h}) = 0$$

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{h}) = \vec{\nabla} (\vec{\nabla} \cdot \vec{h}) - \nabla^2 \vec{h}$$

$$\vec{\nabla} \cdot \vec{\nabla} \vec{h} = \nabla^2 \vec{h} = a\ vector\ field$$

Here is an example of taking a divergence that will be extremely useful. If

$$\vec{r} = x\hat{\imath} + y\hat{} + z\hat{k}$$

and

$$\vec{k} = k_x\hat{\imath} + k_y\hat{} + k_z\hat{k}$$

and

$$\vec{E} = E_x\hat{\imath} + E_y\hat{} + E_z\hat{k}$$

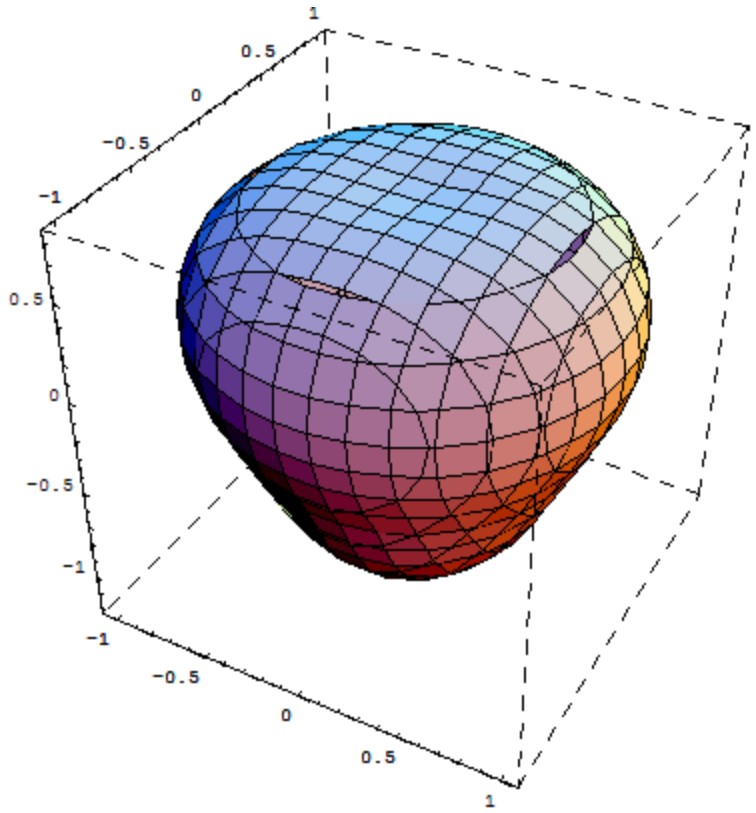
then lets find (for $\vec{E}$ is a constant vector)

$$\begin{aligned}
\vec{\nabla} \cdot \vec{E} e^{i\vec{k}\cdot\vec{r}} &= \tfrac{\partial}{\partial x} E_x e^{i\vec{k}\cdot\vec{r}} + \tfrac{\partial}{\partial y} E_y e^{i\vec{k}\cdot\vec{r}} + \tfrac{\partial}{\partial z} E_z e^{i\vec{k}\cdot\vec{r}} \\
&= E_x e^{i\vec{k}\cdot\vec{r}} \tfrac{\partial}{\partial x} (i\vec{k} \cdot \vec{r}) + \ldots \\
&= iE_x e^{i\vec{k}\cdot\vec{r}} \tfrac{\partial}{\partial x} (k_x x + k_y y + k_z z) + \ldots \\
&= ik_x E_x e^{i\vec{k}\cdot\vec{r}} + ik_y E_y e^{i\vec{k}\cdot\vec{r}} + ik_z E_z e^{i\vec{k}\cdot\vec{r}} \\
&= i\vec{k} \cdot \vec{E} e^{i\vec{k}\cdot\vec{r}}
\end{aligned}$$

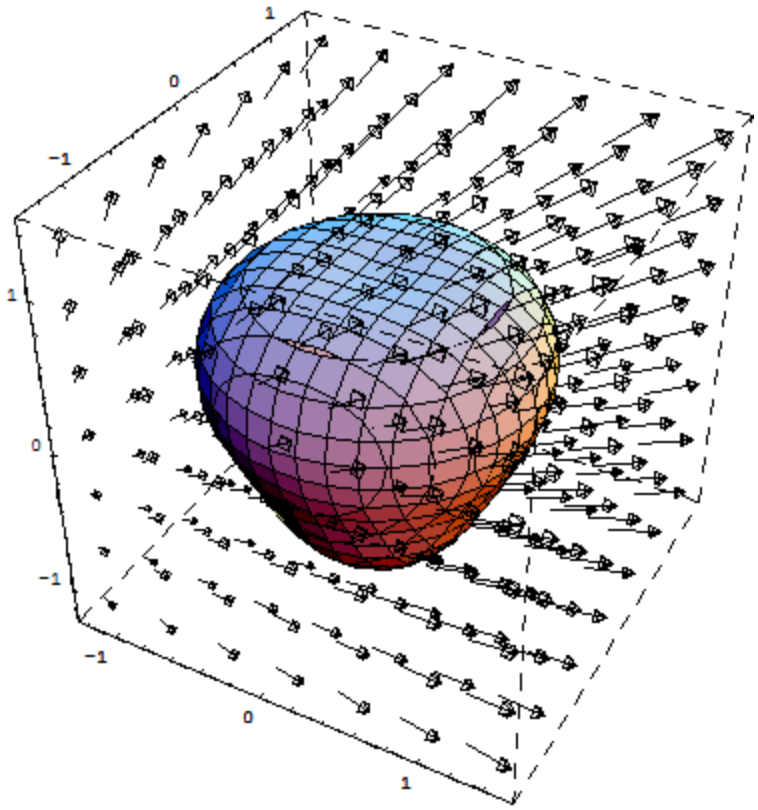
Gauss' Theorem

Gauss' Theorem

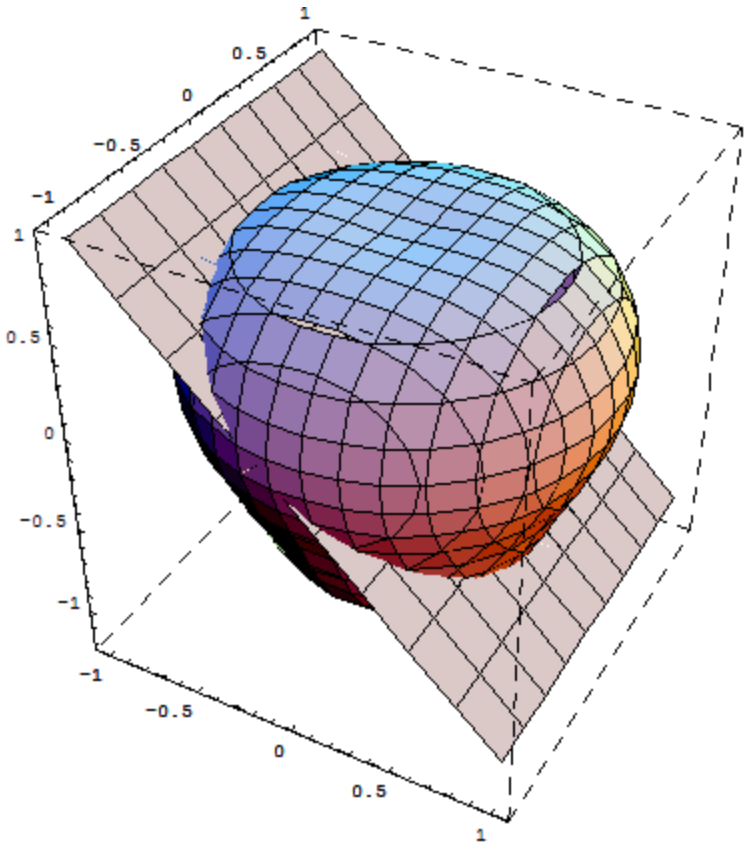
Consider the following volume enclosed by a surface we will call S.



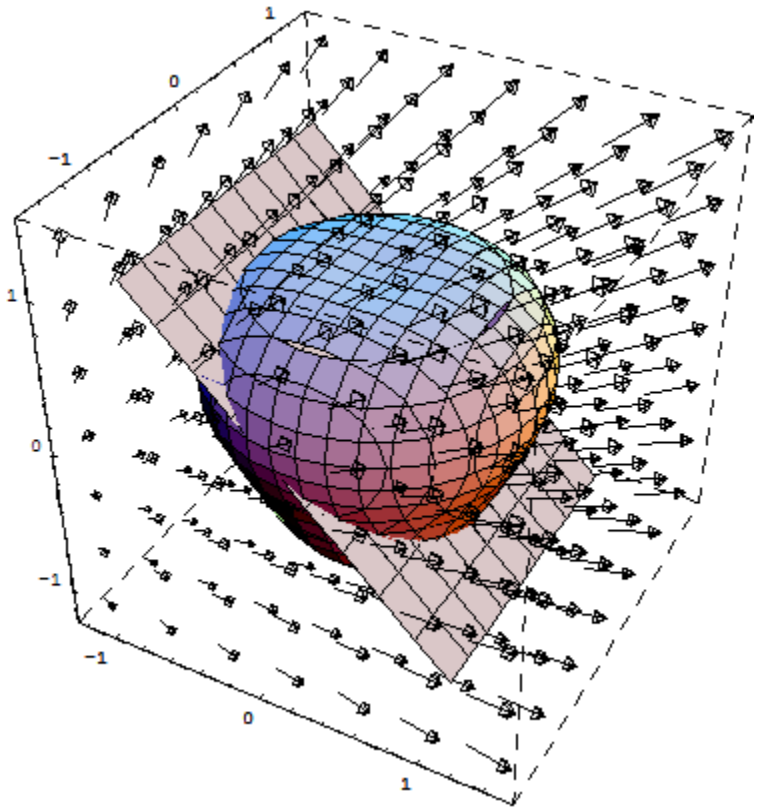
Now we will embed S in a vector field:



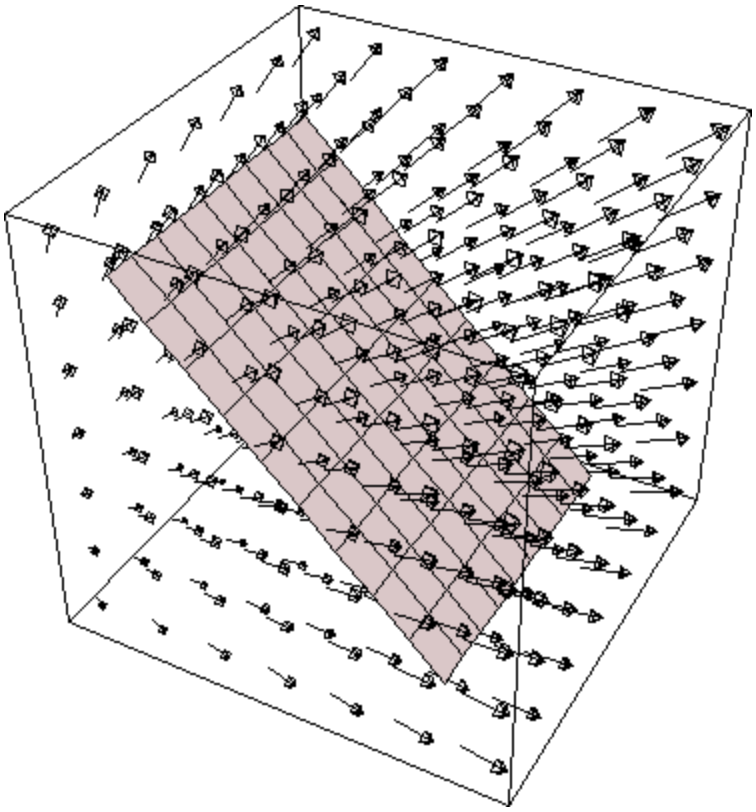
We will cut the the object into two volumes that are enclosed by surfaces we will call S_1 and S_2.



Again we embed it in the same vector field.



It is clear that flux through $S_1 + S_2$ is equal to flux through S.This is because the flux through one side of the plane is exactly opposite to the flux through the other side of the plane:



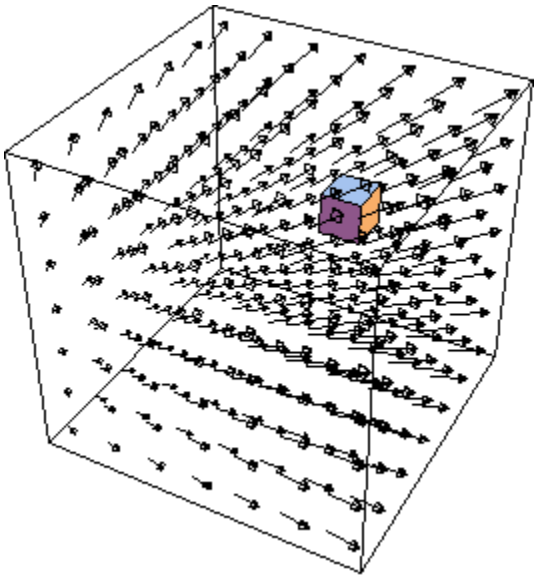
So we see that

$$\oint_S \vec{F} \cdot d\vec{a} = \oint_{S_1} \vec{F} \cdot d\vec{a_1} + \oint_{S_2} \vec{F} \cdot d\vec{a_2}.$$

We could subdivide the surface as much as we want and so for n subdivisions the integral becomes:

$$\oint_S \vec{F} \cdot d\vec{a} = \sum_{i=1}^{n} \oint_{S_i} \vec{F} \cdot d\vec{a_i}.$$

What is $\oint_{S_i} \vec{F} \cdot d\vec{a_i}$.? We can subdivide the volume into a bunch of little cubes:



To first order (which is all that matters since we will take the limit of a small volume) the field at a point at the bottom of the box is

$$F_z + \frac{\Delta x}{2}\frac{\partial F_z}{\partial x} + \frac{\Delta y}{2}\frac{\partial F_z}{\partial y}$$

where we have assumed the middle of the bottom of the box is the point $(x + \frac{\Delta x}{2}, y + \frac{\Delta y}{2}, z)$. Through the top of the box $(x + \frac{\Delta x}{2}, y + \frac{\Delta y}{2}, z + \Delta z)$you get

$$F_z + \frac{\Delta x}{2}\frac{\partial F_z}{\partial x} + \frac{\Delta y}{2}\frac{\partial F_z}{\partial y} + \Delta z\frac{\partial F_z}{\partial z}$$

Through the top and bottom surfaces you get Flux Top - Flux bottom

Which is

$$\Delta x \Delta y \Delta z \frac{\partial F_z}{\partial z} = \Delta V \frac{\partial F_z}{\partial z}$$

Likewise you get the same result in the other dimensionsHence

$$\oint_{S_i} \vec{F} \cdot d\vec{a}_i = \Delta V_i \left[\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \right]$$

or

$$\oint_{S_i} \vec{F} \cdot d\vec{a}_i = \vec{\nabla} \cdot \vec{F} \Delta V_i$$

$$\oint_S \vec{F} \cdot d\vec{a} = \sum_{i=1}^{n} \oint_{S_i} \vec{F} \cdot d\vec{a}_i$$
$$= \sum_{i=1}^{n} \vec{\nabla} \cdot \vec{F} \Delta V_i$$

So in the limit that $\Delta V_i \to 0$ and $n \to \infty$

$$\oint_S \vec{F} \cdot d\vec{a} = \oint_V \vec{\nabla} \cdot \vec{F} dV$$

This result is intimately connected to the fundamental definition of the divergence which is

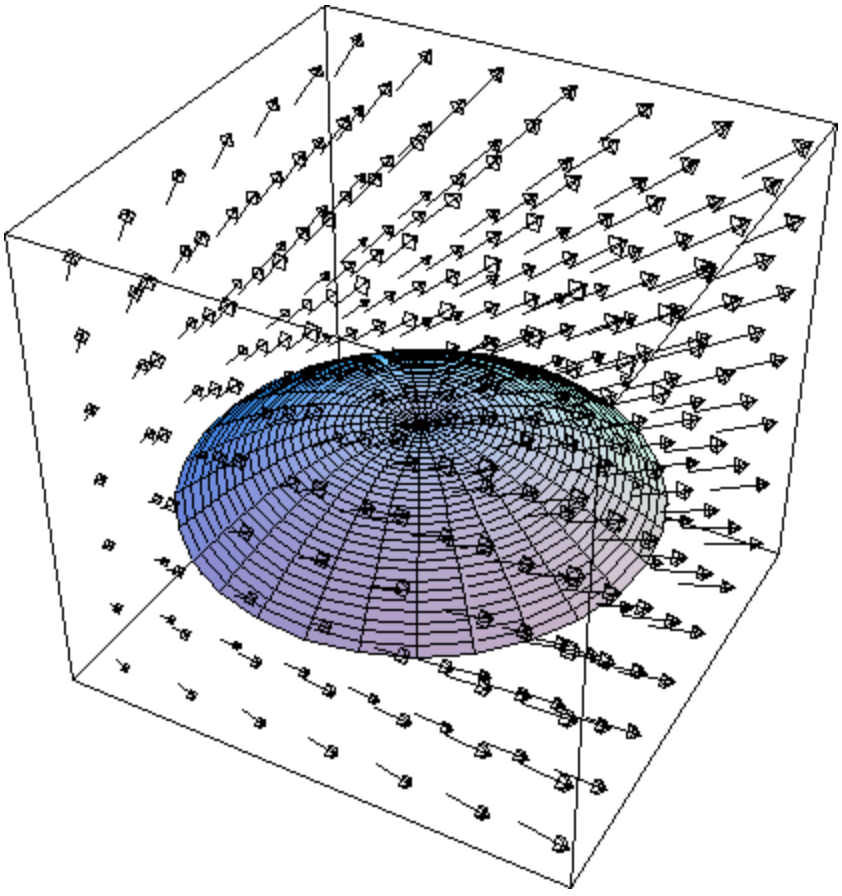
$$\vec{\nabla} \cdot \vec{F} \equiv \lim_{V \to 0} \frac{1}{V} \oint_S \vec{F} \cdot d\vec{a}$$

where the integral is taken over the surface enclosing the volume V. The divergence is the flux out of a volume, per unit volume, in the limit of an infinitely small volume. By our proof of Gauss' theorem, we have shown that the del operator acting on a vector field captures this definition.

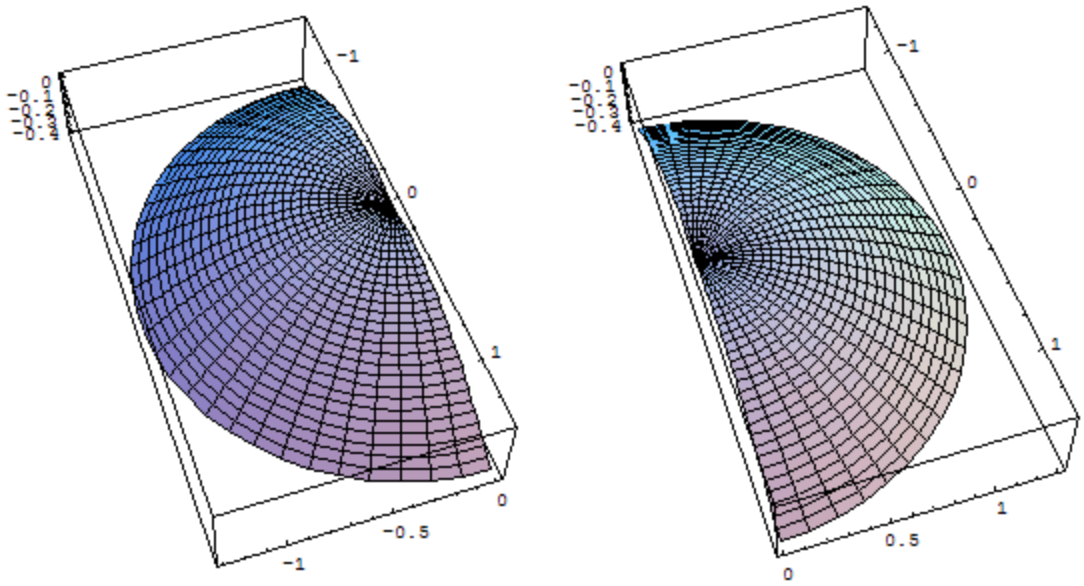
Module 17.xhtml

Stokes' theorem

This is derived in a similar fashion to Gauss' theorem, except now, instead of considering a volume and taking a surface integral, we consider a surface and take a line integral around the edge of the surface. In this case the surface does not enclose a volume. A good picture that describes what is being done is figure 2.23 in Berkeley Physics Course Volume 2. (Unfortunately it is copyrighted, and so it can not be shown on the web - If somebody reading this can provide some suitable drawings I would incorporate them with an acknowledgment). The following figures are not as good as in the book, but will have to do for now. We have some surface with a vector field passing through it:



We can take the line integral around the edge of the surface and evaluate that. We can also slice the surface into two parts



The line integral along the common edge will have opposite signs for each half and so the sum of the two individual line integrals will equal the line integral of the complete surface. We can subdivide as much as we want and we will always have:

$$\oint_C \vec{F} \cdot d\vec{s} = \sum_{i=1}^{N} \oint_{C_i} \vec{F} \cdot d\vec{s}_i.$$

Since we can subdivide as much as we want we can break the surface down into a collection of tiny squares, each of which lies in the $\{x, y\}$, $\{y, z\}$ or $\{x, z\}$ planes.

Consider asquare in the x y plane and lets find the line integral Let us find the line integral of $\vec{F}$. First consider the x component at the center of the bottom of the square

$$F_x = F_x(x,y) + \frac{\Delta x}{2}\frac{\partial F_x}{\partial x}$$

and then at the center of the top

$$F_x = F_x(x,y) + \frac{\Delta x}{2}\frac{\partial F_x}{\partial x} + \Delta y \frac{\partial F_x}{\partial y}$$

So multiplying by Δx and subtracting the top and bottom to get the F_x contribution to the line integral gives

$$-\Delta x \Delta y \frac{\partial F_x}{\partial y}$$

Minus sign comes about from the integration direction (if at the top F_x is more positive this results in a negative contribution)Similarly in y we get the contribution

$$\Delta x \Delta y \frac{\partial F_y}{\partial x}$$

because F_y is more positive in the right you get a positive contribution

So for this simple example

$$\oint_{C_i} \vec{F} \cdot d\vec{s_i} = \Delta x \Delta y \left[\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right]$$

which is just $\left(\vec{\nabla} \times \vec{F}\right)_z$ or if we define $\Delta \vec{A}$ such that it has the area $\Delta x \Delta y$ and a direction perpendicular to the xy plane it is

$$\oint_{C_i} \vec{F} \cdot d\vec{s_i} = \Delta \vec{A} \cdot \left(\vec{\nabla} \times \vec{F}\right)$$

We have shown the above is true for a square in the $\{x, y\}$ plane. Similarly we would look at the other possible planes and in $\{xz\}$ plane would get

$$\Delta x \Delta z \left[\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right]$$

which is $\left(\vec{\nabla} \times \vec{F} \right)_y$. In the $\{yz\}$ plane we get

$$\Delta y \Delta z \left[\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right]$$

which is $\left(\vec{\nabla} \times \vec{F} \right)_x$

So in the 3d case we see that Now

$$\oint_C \vec{F} \cdot d\vec{s} = \sum_{i=1}^{N} \oint_{C_i} \vec{F} \cdot d\vec{s}_i$$

becomes

$$\oint_C \vec{F} \cdot d\vec{s} = \sum_{i=1}^{N} \Delta \vec{A}_i \cdot \vec{\nabla} \times \vec{F}$$

which in the infinitesimal limit is

$$\oint_C \vec{F} \cdot d\vec{s} = \int_S d\vec{A} \cdot \vec{\nabla} \times \vec{F}.$$

This is the fundamental definition of curl. We have shown that the del operator "crossed" into the vector field captures the definition of the operation.

As a side note, Physicists uniformly refer to the above as Stoke's theorem but in fact that is a more general theorem and the above can correctly be called the "curl theorem". This is a physics course though, so we will call it Stoke's theorem.

Module 18.xhtml

Gauss' Law

Now recall that flux is the scalar product of a vector field and a bit of surface

$$Flux = \vec{F} \cdot \vec{a}$$

where $\vec{F}$ is some vector field and $\vec{a}$ is a surface with the direction defined by the normal to the surface. For a series of connected surfaces $\vec{a}_j$ the total flux through the combined surface would be the sum of the individual elements. For a vector field $\vec{E}$ passing through the surface this leads to

$$\Phi = \sum \vec{E}_j \cdot \vec{a}_j$$

or when we go to infinitesimal areas

$$\Phi = \int_{surface} \vec{E} \cdot d\vec{a}$$

Now lets consider a charge *q* in the middle of a sphere

$$\begin{aligned} \Phi = \oint \vec{E} \cdot d\vec{a} &= \oint E \cdot \mathbb{d}a \\ &= E \oint \mathbb{d}a \\ &= E(4\pi r^2) \end{aligned}$$

but

$$E = \frac{kq}{r^2}$$

then

$$\Phi = 4\pi kq$$

$$\varepsilon_0 \equiv \frac{1}{4\pi k}$$

So for this case we get

$$\oint \vec{E} \cdot d\vec{a} = \frac{q}{\varepsilon_0}$$

We can generalize this to any closed surface. It is clear that for an arbitrary closed source, we can draw a sphere around the source within the arbitrary surface.. Think of bullets being fired from a gun, it is clear that the bullets originating in the inner sphere all pass through the outer surface and so one would expect that the flux would be the same. For example consider $\vec{a}$ to be a patch on the inner sphere and $\vec{A}$ to be its projection onto the outer arbitrary surface (with its normal making an angle θ with respect to the normal to $\vec{a}$.

On the inner patch

$$\Phi_r = \vec{E}_r \cdot \vec{a} = E_r a$$

and at the outer patch

$$\begin{aligned}
\Phi_R &= \vec{E}_R \cdot \vec{A} \\
&= E_R A \cos\theta \\
&= \left[E_r \left(\tfrac{r}{R}\right)^2\right] \left[a \left(\tfrac{R}{r}\right)^2 \tfrac{1}{\cos\theta}\right] \cos\theta \\
&= E_r a \\
&= \Phi_r
\end{aligned}$$

So the two have equivalent fluxes.

Any electric field is the sum of fields of its individual sources so we can write

$$\Phi = \oint \vec{E} \cdot d\vec{a} = \oint \sum_i \vec{E}_i d\vec{a}$$
$$= \frac{1}{\varepsilon_0} \sum_i q_i$$

or for charge distributed throughout the volume

$$\oint \vec{E} \cdot d\vec{a} = \frac{1}{\varepsilon_0} \int \rho \mathbb{d}V$$

Now we can apply Gauss' Theorem

$$\oint \vec{E} \cdot d\vec{a} = \int \vec{\nabla} \cdot \vec{E} \mathbb{d}V = \frac{1}{\varepsilon_0} \int \rho \mathbb{d}V$$

The equation

$$\int \vec{\nabla} \cdot \vec{E} \mathbb{d}V = \frac{1}{\varepsilon_0} \int \rho \mathbb{d}V$$

must be true for any volume of any size, shape or location. The only way that can be true is if:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon_0}$$

Initially one may think that this is a much less clear way of posing Gauss' Law. In practice it is much more useful than the integral form. Given an arbitrary distribution of charge we can calculate the electric field anywhere in space.

Gauss' Law for Magnetism

We can consider the same arguments for magnetic fields however there is one major difference! There are no isolated source of magnetism. That is there are no magnetic monopoles. This is an experimental fact. In fact people continue to search for them but they have never been found. (Finding one would almost certainly be a discovery worthy of a Nobel Prize). So we have

$$\oint \vec{B} \cdot d\vec{a} = 0$$

or

$$\vec{\nabla} \cdot \vec{B} = 0.$$

Module 19.xhtml

Faraday's Law

The magnetic flux is

$$\Phi_M = \int \vec{B} \cdot d\vec{A}$$

A device that can maintain a potential difference, despite the flow of current is a source of electromotive force. (EMF) The definition is mathematically

$$\varepsilon = \oint \vec{E} \cdot d\vec{s}$$

Faraday's law states that

$$\varepsilon = -\frac{\partial \Phi_M}{\partial t}$$

This is simply an experimental fact.

Faraday's law could also be written

$$\oint \vec{E} \cdot d\vec{s} = -\frac{\partial}{\partial t} \int \vec{B} \cdot d\vec{A}$$

I will leave it as a problem to show that this can also be written

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}.$$

Module 20.xhtml

Ampere's Law (with displacement current)

For a steady current flowing through a straight wire, the magnetic field at a point at a perpendicular distance r from the wire, has a value

$$B = \frac{\mu_0 I}{2\pi r}$$

If we integrate around the wire in a circle, then clearly we get

$$\begin{aligned} \oint \vec{B} \cdot d\vec{l} &= \tfrac{\mu_0 I}{2\pi r} \oint \mathrm{d}l \\ &= \tfrac{\mu_0 I}{2\pi r} 2\pi r \\ &= \mu_0 I \end{aligned}$$

This is true for irregular paths around the wire

$$\oint \vec{B} \cdot d\vec{l} = \oint B \cdot \mathrm{d}l \cos\theta$$

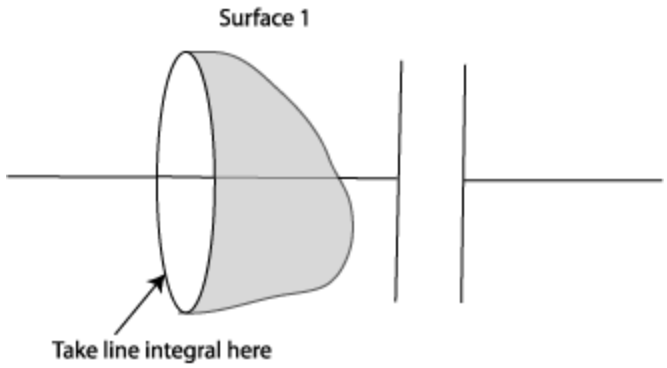
but for small dl $dl \cos\theta = r d\varphi$

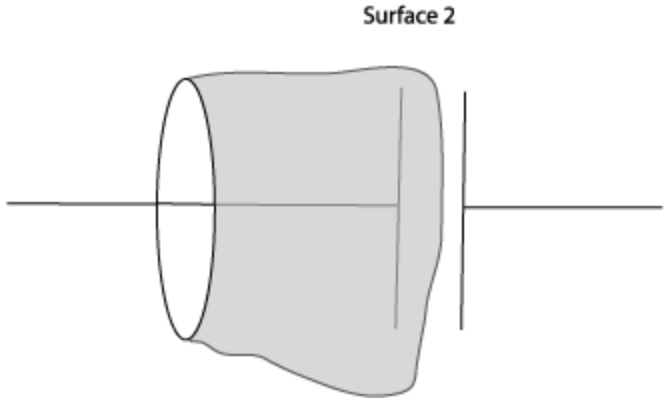
$$\begin{aligned} \oint \vec{B} \cdot d\vec{l} &= \oint Br\mathrm{d}\varphi \\ &= \tfrac{\mu_0 I}{2\pi} \oint \mathrm{d}\varphi \\ &= \mu_0 I \end{aligned}$$

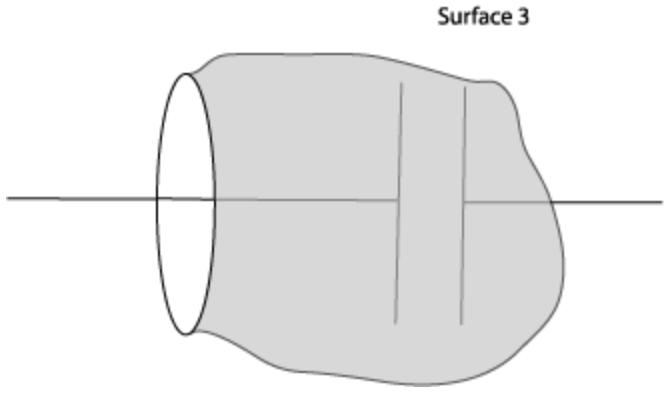
In fact instead of current we use the surface integral of the current density J , which is the current per unit area

$$\oint \vec{B} \cdot d\vec{l} = \mu_0 \int_S \vec{J} \cdot d\vec{A}$$

Maxwell's great insight was to realize that this was incomplete. He reasoned that $\frac{\mathrm{d}\varphi_B}{\mathrm{d}t}$ gives a $\vec{E}$ field so we should expect that $\frac{\mathrm{d}\varphi_E}{\mathrm{d}t}$ gives a $\vec{B}$ field.







Think of a capacitor in a simple circuit. We can draw a surface such as shown in the figure, with "surface 1" and take the line integral around the edge of the surface. Now look at surface 2, this will have the same line integral, but now the surface integral will be different. Clearly there is something incomplete with Ampere's law as formulated above. Maxwell re wrote Ampere's law

$$\oint \vec{B} \cdot d\vec{l} = \mu_0 \int_S \left(\vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \right) \cdot d\vec{A}$$

which solves the problem.

Again it is left as an exercise to show that

$$\vec{\nabla} \times \vec{B} = \mu_0 \left(\vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \right)$$

Maxwell's equations

Lets recall Maxwell's equations (in free space) in differential form

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$
$$\vec{\nabla} \times \vec{B} = \mu_0 \left(\vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \right)$$
$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon_0}$$
$$\vec{\nabla} \cdot \vec{B} = 0$$

Module 21.xhtml

The E&M Wave Equation

Lets recall Maxwell's equations in differential form

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$
$$\vec{\nabla} \times \vec{B} = \mu_0 \left(\vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \right)$$
$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon_0}$$
$$\vec{\nabla} \cdot \vec{B} = 0$$

In free space there are no charges or currents these become:

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$
$$\vec{\nabla} \times \vec{B} = \mu_0 \varepsilon_0 \frac{\partial \vec{E}}{\partial t}$$
$$\vec{\nabla} \cdot \vec{E} = 0$$
$$\vec{\nabla} \cdot \vec{B} = 0$$

Lets take the time derivative of

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

$$\vec{\nabla} \times \frac{\partial \vec{E}}{\partial t} = -\frac{\partial^2 \vec{B}}{\partial t^2}$$

$$\vec{\nabla} \times \vec{\nabla} \times \vec{B} = -\mu_0 \varepsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2}$$

but recall

$$\vec{\nabla} \times \vec{\nabla} \times \vec{C} = \vec{\nabla}(\vec{\nabla} \cdot \vec{C}) - (\vec{\nabla} \cdot \vec{\nabla})\vec{C}$$

so using that and $\vec{\nabla} \cdot \vec{B} = 0$ we get

$$\nabla^2 \vec{B} = \mu_0 \varepsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2}$$

This is the 3d wave equation! Note that is a second time derivative on one side and a second space derivative on the other side It is left as an exercise to show that

$$\nabla^2 \vec{E} = \mu_0 \varepsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}$$

we also see from this equation that the speed of light in vacuum is

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

module 22.xhtml

Plane Waves

We want to find the expression for a plane that is perpendicular to $\vec{k}$, where $\vec{k}$ is a vector in the direction of propagation of the wave.The plane is the set of points that has the same projection onto the vector $\vec{k}$ That is any point $\vec{r}$ that satisfies

$$\vec{k} \cdot \vec{r} = constant$$

is a point on the planeNow consider the function

$$\psi(\vec{r}) = Ae^{i\vec{k}\cdot\vec{r}}$$

we see that the magnitude of $\psi(\vec{r})$ is the same over every plane that is defined by $\vec{k} \cdot \vec{r} = constant$ we want to construct harmonic waves, ie. they should repeat every wavelength along the direction of propagation so they should satisfy

$$\psi(\vec{r}) = \psi\left(\vec{r} + \frac{\lambda\vec{k}}{k}\right)$$

where λ is the wavelengththen we must have

$$\begin{aligned} Ae^{i\vec{k}\cdot\vec{r}} &= Ae^{i\vec{k}\cdot\left(\vec{r}+\frac{\lambda\vec{k}}{k}\right)} \\ &= Ae^{i\vec{k}\cdot\vec{r}}e^{i\vec{k}\cdot\vec{k}\lambda/k} \\ &= Ae^{i\vec{k}\cdot\vec{r}}e^{ik\lambda} \end{aligned}$$

This is true if

$$e^{i\lambda k} = 1 = e^{i2\pi}$$

or

$$\lambda k = 2\pi$$

$$k = \frac{2\pi}{\lambda}$$

This should have a familiar look to it! Finally we want these waves to propagate in time so you should be able to guess the answer from our work on mechanical waves

$$\psi(r) = Ae^{i\left(\vec{k}\cdot\vec{r}\mp\omega t\right)}$$

Spherical Waves

Spherical Waves

To find spherical solutions to the wave equation it is natural to use spherical coordinates.

$$x = r \sin\theta \cos\varphi$$

$$y = r \sin\theta \sin\varphi$$

$$z = r \cos\theta$$

There is a very nice discussion of Spherical Coordinates at:

http://mathworld.wolfram.com/SphericalCoordinates.html

There is also a nice discussion of Cylindrical coordinates at the same site

http://mathworld.wolfram.com/CylindricalCoordinates.html

Beware the confusion about θ and φ. We are calling the polar angle θ. All other mathematical disciplines get it wrong and call it φ.

The Laplacian can be written in spherical coordinates, but where does that come from?looking at just the x term

$$\frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x}\frac{\partial}{\partial \theta} + \frac{\partial \varphi}{\partial x}\frac{\partial}{\partial \varphi}$$

Then you take the second derivative to get $\frac{\partial^2}{\partial x^2}$ which as you can imagine is a tremendously boring and tedious thing to do.Since this isn't a vector calculus course lets just accept the solution.In the case of spherical waves it is not so difficult since the θ and φ derivative terms all go to 0.

$$\begin{aligned}\nabla^2\psi(r) &= \tfrac{1}{r^2}\tfrac{\partial}{\partial r}\left(r^2\tfrac{\partial\psi}{\partial r}\right)\\ &= \tfrac{\partial^2\psi}{\partial r^2}+\tfrac{2}{r}\tfrac{\partial\psi}{\partial r}\\ &= \tfrac{1}{r}\tfrac{\partial}{\partial r}\left(\psi+r\tfrac{\partial\psi}{\partial r}\right)\\ &= \tfrac{1}{r}\tfrac{\partial^2(r\psi)}{\partial r^2}\end{aligned}$$

Thus for spherical waves, we can write the wave equation:

$$\frac{1}{r}\frac{\partial^2\left(r\psi\right)}{\partial r^2}=\frac{1}{\mathrm{v}^2}\frac{\partial^2\psi}{\partial t^2}$$

Now we can multiply both sides by r and since r does not depend upon t write

$$\frac{\partial^2\left(r\psi\right)}{\partial r^2}=\frac{1}{\mathrm{v}^2}\frac{\partial^2}{\partial t^2}(r\psi)$$

This is just the one dimensional wave equation with a harmonic solution

$$r\psi(r,t)=Ae^{ik(r\mp\mathrm{v}t)}$$

or

$$\psi(r,t)=\frac{A}{r}e^{ik(r\mp\mathrm{v}t)}$$

module 24.xhtml

Transverse Waves

A plane wave solution to the electromagnetic wave equation for the $\vec{E}$ field is

$$\vec{E}(\vec{r},t) = \vec{E}_0 e^{i\left(\vec{k}\cdot\vec{r}-\omega t\right)}$$

In vacuum with no currents present we know that: $\vec{\nabla}\cdot\vec{E} = 0$. Recall that earlier we showed

$$\vec{\nabla}\cdot\vec{E} = i\vec{k}\cdot\vec{E}_0 e^{i\left(\vec{k}\cdot\vec{r}-\omega t\right)}$$

So

$$\vec{\nabla}\cdot\vec{E} = i\vec{k}\cdot\vec{E}_0 e^{i\left(\vec{k}\cdot\vec{r}-\omega t\right)} = 0$$

implies that the $\vec{E}$ associated with our plane wave is perpendicular to its direction of motion.

Likewise $\vec{\nabla}\cdot\vec{B} = 0$ implies that the $\vec{B}$ field is also perpendicular to the direction of motion Lets pick a specific simple case:

$$\vec{E} = \hat{}E_y(x,t)$$

Then Faraday's law

$$\vec{\nabla}\times\vec{E} = -\frac{\partial\vec{B}}{\partial t}$$

tells us that (since $\frac{\partial E_y}{\partial z} = 0$)

$$\frac{\partial E_y}{\partial x}\widehat{k} = -\frac{\partial B_z}{\partial t}\widehat{k}$$

That is the $\vec{B}$ field is at Right angles to the $\vec{E}$ field.Also

$$
\begin{aligned}
B_z &= -\int \frac{\partial E_y}{\partial x} \mathrm{d}t \\
&= -\int \frac{\partial}{\partial x} E_0 e^{i(kx-\omega t)} \mathrm{d}t \\
&= -\int \frac{\partial}{\partial x} E_0 e^{ikx} e^{-i\omega t} \mathrm{d}t \\
&= -\frac{\partial}{\partial x} E_0 e^{ikx} \int e^{-i\omega t} \mathrm{d}t \\
&= -ikE_0 e^{ikx} \int e^{-i\omega t} \mathrm{d}t \\
&= -ikE_0 e^{ikx} \frac{e^{-i\omega t}}{-i\omega} \\
&= \frac{1}{c} E_0 e^{ikx} e^{-i\omega t} \\
&= \frac{1}{c} E_y
\end{aligned}
$$

I leave as an exercise showing $\frac{k}{\omega} = \frac{1}{c}$

A movie demonstrating a plane wave can be seen at

http://www.cs.brown.edu/stc/outrea/greenhouse/nursery/physics/gfx/emwave.mov

An applet can be viewed at

http://www.phy.ntnu.edu.tw/java/emWave/emWave.html

Module 25.xhtml

Energy Density

The electric and magnetic fields have energy and hence have an energy density. We can see this for a capacitor:The energy stored in a capacitor is

$$U = \frac{1}{2}CV^2$$

where C is the capacitance and V the potential drop (voltage) across the capacitor. For a parallel plate capacitor $C = \frac{\varepsilon_0 A}{d}$ and $V = Ed$ where A is the area of the plates d the distance between them and E the electric field strength.note that Ad is the volumeThus

$$U = \frac{1}{2}\frac{\varepsilon_0 A}{d}(Ed)^2 = \frac{1}{2}\varepsilon_0 AdE^2$$

So we can write the energy density (Energy per Unit volume) of the field as

$$u_E = \frac{U}{Ad} = \frac{1}{2}\varepsilon_0 E^2$$

Likewise by calculating the energy stored by a B-field in a current carrying solenoid one can derive:

$$u_B = \frac{B^2}{2\mu_0}$$

Since we know $E = cB$

$$
\begin{aligned}
u_E &= \tfrac{1}{2}\varepsilon_0 E^2 \\
&= \tfrac{1}{2}\varepsilon_0 c^2 B^2 \\
&= \tfrac{1}{2}\varepsilon_0 \tfrac{1}{\varepsilon_0 \mu_0} B^2 \\
&= \tfrac{1}{2}\tfrac{1}{\mu_0} B^2 \\
&= u_B
\end{aligned}
$$

In an EM wave $u = u_E + u_B$ which is $u = \varepsilon_0 E^2$ or equivalently $u = B^2/\mu_0$

Module 26.xhtml

Review

Lets take some time to review what we have learned so far. We have derived Maxwell's equations in differential form.

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$
$$\vec{\nabla} \times \vec{B} = \mu_0 \left(\vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \right)$$
$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon_0}$$
$$\vec{\nabla} \cdot \vec{B} = 0$$

These, in general are much more useful than the integral form you learned in Freshman Physics. These allow one to understand the relationship between fields, charges and currents as a function of position. This point by point understanding of what is happening is not obvious in the integral form of the equations.

Another interesting point is that if everything is static, that is nothing is changing with time, then they become

$$\vec{\nabla} \times \vec{E} = 0$$
$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{J}$$
$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon_0}$$
$$\vec{\nabla} \cdot \vec{B} = 0$$

Notice that for static fields, there is no interplay between electricity and magnetism. If there was just electrostatics, then we would have separate electric and magnetic fields. Maxwell was able to show that the electricity and magnetism are intimately related, and the theory is unified in that you need both. (To this day the unification of forces is one of the driving

principles of a lot of physics research - I would say the only interesting physics research but that is perhaps because I do it for a living.)

In free space Maxwell's equations become:

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$
$$\vec{\nabla} \times \vec{B} = \mu_0 \varepsilon_0 \frac{\partial \vec{E}}{\partial t}$$
$$\vec{\nabla} \cdot \vec{E} = 0$$
$$\vec{\nabla} \cdot \vec{B} = 0$$

We then showed that one can take time derivatives and end up with

$$\nabla^2 \vec{B} = \mu_0 \varepsilon_0 \frac{\partial^2 \vec{B}}{\partial t^2}$$

which is the 3d wave equation! Note that is a second time derivative on one side and a second space derivative on the other side, the hallmarks of a wave equation.

It was left as an exercise to show that

$$\nabla^2 \vec{E} = \mu_0 \varepsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2}$$

We also see from this equation that the speed of light in vacuum is

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

A plane wave solution to the electromagnetic wave equation for the $\vec{E}$ field is

$$\vec{E}(\vec{r}, t) = \vec{E}_0 e^{i\left(\vec{k} \cdot \vec{r} - \omega t\right)}$$

In vacuum with no currents present we know that: $\vec{\nabla} \cdot \vec{E} = 0$. Recall that earlier we showed

$$\vec{\nabla} \cdot \vec{E} = i\vec{k} \cdot \vec{E}_0 e^{i\left(\vec{k}\cdot\vec{r} - \omega t\right)}$$

So

$$\vec{\nabla} \cdot \vec{E} = i\vec{k} \cdot \vec{E}_0 e^{i\left(\vec{k}\cdot\vec{r} - \omega t\right)} = 0$$

implies that the $\vec{E}$ associated with our plane wave is perpendicular to its direction of motion.

Likewise $\vec{\nabla} \cdot \vec{B} = 0$ implies that the $\vec{B}$ field is also perpendicular to the direction of motion

The electric and magnetic fields have energy and hence have an energy density. In an EM wave $u = u_E + u_B$ which is $u = \varepsilon_0 E^2$ or equivalently $u = B^2/\mu_0$.This is all very amazing when you think about it. Maxwell's equations tell us that we can have waves in the electric and magnetic fields. These waves carry energy. That is they are a mechanism to transport energy through free space (or a medium). This is why the sun warms us, which is pretty important.

Module 27.xhtml

Poynting Vector

Now we want to calculate the power crossing a given area A. During a time Δt an EM wave will pass an amount of energy through A of $uc\Delta tA$ where u is the energy density of the wave. If we want the power/m^2 then we must divide by ΔtA. Thus we get

$$\begin{aligned} S &= \tfrac{uc\Delta tA}{\Delta tA} \\ &= uc \\ &= \tfrac{1}{\mu_0}B^2c \\ &= \tfrac{1}{\mu_0}B\tfrac{E}{c}c \\ &= \tfrac{1}{\mu_0}BE \end{aligned}$$

Now we make the reasonable assumption that the energy flows in the direction of the wave, ie. perpendicular to $\vec{E}$ and $\vec{B}$ so we can define a vector that has the power per unit area:

$$\vec{S} = \frac{1}{\mu_0}\vec{E} \times \vec{B}$$

or

$$\vec{S} = c^2\varepsilon_0\vec{E} \times \vec{B}$$

This is the Poynting vector. Thus for a plane EM wave we have three useful things

$$\vec{E}(\vec{r},t) = \vec{E}_0 e^{i\left(\vec{k}\cdot\vec{r}-\omega t\right)}$$

$$\vec{B}(\vec{r},t) = \vec{B}_0 e^{i\left(\vec{k}\cdot\vec{r}-\omega t\right)}$$

and

$$\vec{S} = c^2 \varepsilon_0 \vec{E}_0 \times \vec{B}_0 e^{i2\left(\vec{k}\cdot\vec{r} - \omega t\right)}$$

Module 28.xhtml

Irradiance

The frequency of optical light is $\approx 10^{15}$Hz and so the Poynting vector varies extremely quickly. So it is useful to determine an time averaged quantity. So lets define the irradiance as

$$I = \langle S \rangle_T$$

where the symbol $\langle\ \rangle_T$ means find the average over a time T. For T we want to use an integer multiple of periods (such as 1). (What would you end up with otherwise? It wouldn't really make sense to me) This is a case where it is easier to use a trig function for the wave. Let's consider the wave

$$\vec{E}(\vec{r},t) = \vec{E}_0 \cos\ \left(\vec{k}\cdot\vec{r} - \omega t\right).$$

Then

$$S = c^2\varepsilon_0 E_0 B_0 \cos^2\ \left(\vec{k}\cdot\vec{r} - \omega t\right)$$

Now we need to find

$$\langle S \rangle_T = c^2\varepsilon_0 E_0 B_0 \left\langle \cos^2\ \left(\vec{k}\cdot\vec{r} - \omega t\right)\right\rangle_T$$

$$\langle S \rangle_T = c^2\varepsilon_0 E_0 B_0 \frac{1}{T}\int_0^T \cos^2\ \left(\vec{k}\cdot\vec{r} - \omega t\right)\mathrm{d}t$$

we make a coordinate transformation $x = \vec{k}\cdot\vec{r} - \omega t$. Then

$$\langle S\rangle_T = c^2\varepsilon_0 E_0 B_0 \tfrac{1}{T}\int_0^T \cos^2\left(\vec{k}\cdot\vec{r}-\omega t\right)\mathrm{d}t$$

$$= c^2\varepsilon_0 E_0 B_0 \tfrac{1}{T}\int_{\vec{k}\cdot\vec{r}}^{\vec{k}\cdot\vec{r}-\omega T} \cos^2 x \mathrm{d}x \tfrac{\mathrm{d}t}{\mathrm{d}x}$$

$$= c^2\varepsilon_0 E_0 B_0 \tfrac{-1}{\omega T}\int_{\vec{k}\cdot\vec{r}}^{\vec{k}\cdot\vec{r}-\omega T} \cos^2 x \mathrm{d}x$$

$$= c^2\varepsilon_0 E_0 B_0 \tfrac{-1}{\omega T}\int_{\vec{k}\cdot\vec{r}}^{\vec{k}\cdot\vec{r}-\omega T} \left[\tfrac{1+\cos 2x}{2}\right]\mathrm{d}x$$

$$= c^2\varepsilon_0 E_0 B_0 \tfrac{-1}{\omega T}\left[\tfrac{x}{2} + \tfrac{\sin 2x}{4}\right|_{\vec{k}\cdot\vec{r}}^{\vec{k}\cdot\vec{r}-\omega T}$$

$$= c^2\varepsilon_0 E_0 B_0 \tfrac{-1}{\omega T}\tfrac{\vec{k}\cdot\vec{r}-\omega T}{2} - \tfrac{\vec{k}\cdot\vec{r}}{2} + \tfrac{\sin 2\left(\vec{k}\cdot\vec{r}-\omega T\right)}{4} - \tfrac{\sin 2\left(\vec{k}\cdot\vec{r}\right)}{4}$$

$$= c^2\varepsilon_0 E_0 B_0 \tfrac{-1}{\omega T}\left[\tfrac{-\omega T}{2}\right]$$

$$= c^2\varepsilon_0 E_0 B_0/2$$

Note that above I assume that I can pick T to be an integer multiple of the period.

Now we have just gone through a rather involved derivation of something that should of been intuitively obvious to us. The time average of a harmonic function is zero. The time average of its square is 1/2. Small digression. For a harmonic function, a useful quantity is the Root Mean Square: RMS. For example look at the power rating on a stereo, they have to specify whether it is peak, or RMS power that they are referring to.

Momentum and Radiation Pressure

Remember that $< S >$ is the average power per unit area in the wave. So the Power in the wave is $< S > A$ where A is the area crossed. The force of the wave when it hits something is Power/velocity or

$$F = \frac{< S > A}{c}$$

So the electromagnetic wave can exert a pressure (on a black object)

$$pressure = \frac{F}{A} = \frac{< S >}{c}$$

If it hits a reflective surface then it is

$$pressure = 2\frac{< S >}{c}$$

Also we can write:

$$F = \frac{\mathbb{d}p}{\mathbb{d}t} = \frac{power}{c} = \frac{1}{c}\frac{\mathbb{d}E}{\mathbb{d}t}$$

(here E is the energy). So the momentum in a unit volume of the EM field is the Energy in unit volume of the EM field/c. The direction of the momentum is the direction of propagation of the wave. That is:

$$u = \frac{B^2}{\mu_0} = \frac{EB}{c\mu_0} = \frac{S}{c}$$

and

$$momentum/unit\ volume \equiv \vec{g} = \frac{u}{c}\widehat{k} = \frac{\vec{S}}{c^2}$$

Radiation

A stationary charge or a uniformly moving charge can not produce an EM wave (or radiation). This is obvious when you consider a stationary charge. You would see a time independent $\vec{E}$ field around it but no $\vec{B}$ field. Thus there would not be a Poynting vector and no photons would be emitted.

What if you were driving by the charge at a constant speed. Then you would measure an $\vec{E}$ and $\vec{B}$ field but the irradience would integrate to zero. If you stopped moving with respect to the charge this can't make photons appear or disappear. The photons don't know what you are doing! If a charge moves nonuniformly though it will radiate.

Suppose you have an oscillating dipole $\vec{\wp} = \vec{\wp}_0 \cos \omega t$ When you get far from the dipole you get a wave with a fixed wavelength

$$E = \frac{\wp_0 k^2 \sin\theta}{4\pi\varepsilon_0} \frac{\cos\ (kr - \omega t)}{r}$$

Here E is the electric field intensity, θ is with respect to the dipole moment (see figure 3.33 in Hecht Fourth Edition or figure 3.31 in the Third edition).The irradiance from this is given by

$$< S >_T = I(\theta) = \frac{\wp_0^2 \omega^4}{32\pi^2 c^3 \varepsilon_0} \frac{\sin^2\theta}{r^2}$$

one can integrate over the angle (at any radius) and get the total energy radiated

$$\int < S >_T \, \mathrm{d}\Omega = \frac{\wp_0^2 \omega^4}{12\pi c^3 \varepsilon_0}$$

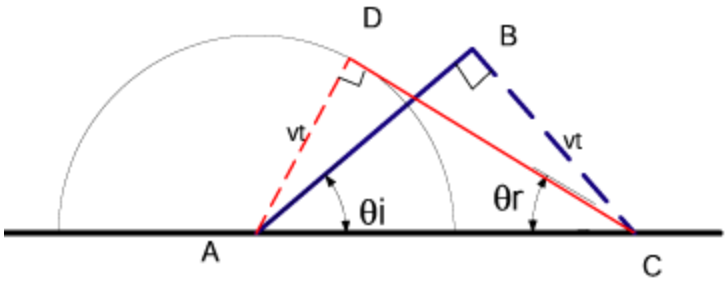
Module 29.xhtml

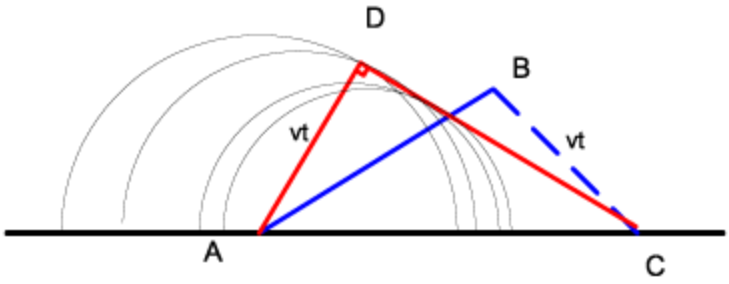
The Huygens-Fresnel Principle

In order to proceed with the discussion we have to define two terms. A wave front is the surface of constant phase. In a plane wave these are planes and in a spherical wave these are spheres. A ray travels perpendicular to the fronts.

Huygens postulated that as a wave propagates through a medium each point on the advancing wavefront acts as a new point source of the wave. This is correct physics for the water waves but not for light waves. However the Helmholtz equation for diffraction of EM waves gives a solution identical to that give by Huygens' principle.

Look at the figure which shows a wavefront AB coming to a surface and is reflected creating the front CD. The point A hits the surface first. The point B hits a time vt later. During that time a spherical wave is emitted from A and travels a distance vt. In fact this happens for every point along the wavefront. The next figure attempts to show how a number of waves line up along the line DC and that this is perpendicular to the line AD.





From this we see that

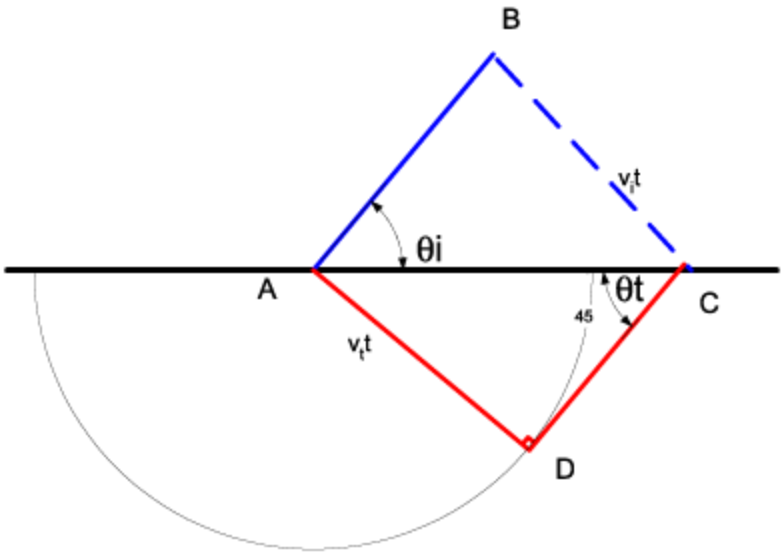
$$\sin\theta_i = \frac{vt}{AC}$$

and

$$\sin\theta_r = \frac{vt}{AC}$$

so $\theta_i = \theta_r$

For refraction a similar thing happens. See figure (geometric optics / Huygens refraction.vsd)



In this case the velocities are different in the two media and so one obtains:

$$\sin\theta_i = \frac{v_i t}{AC}$$

and

$$\sin\theta_t = \frac{v_t t}{AC}$$

which then can be rearranged

$$\frac{\sin\theta_i}{v_i t} = \frac{\sin\theta_t}{v_t t}$$

or rearranging some more

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{v_i t}{v_t t}$$

or

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{n_t}{n_i}$$

finally

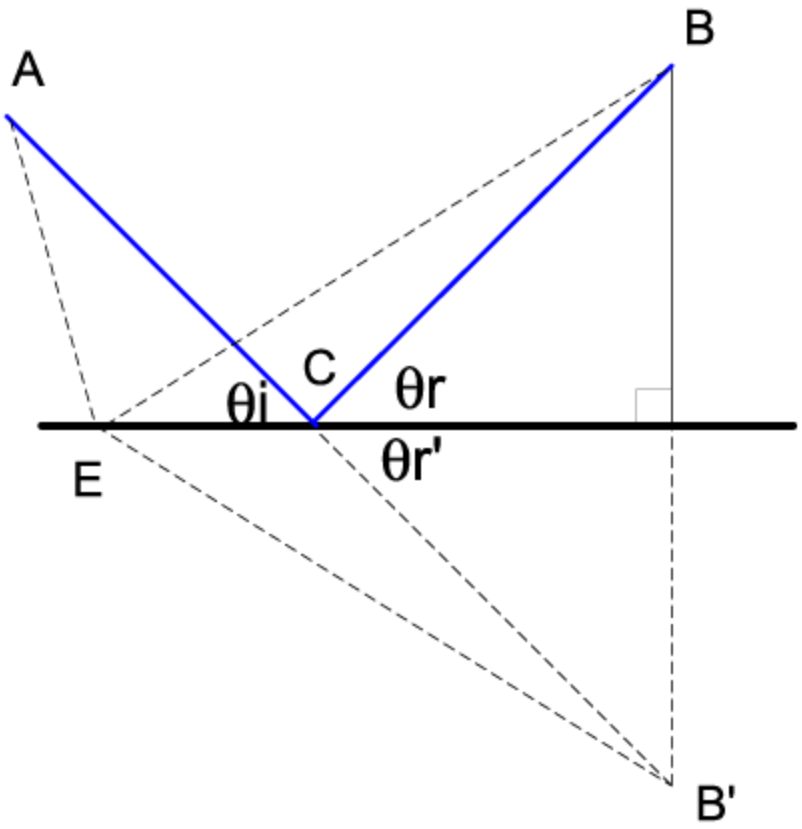
$$n_t \sin \theta_t = n_i \sin \theta_i$$

which is Snell's law. Now note that normally one uses rays, in which case the angles are measured w.r.t. the normal to the surface.

Module 30.xhtml

Fermat's principle of least time

Fermat postulated that rays of light follow the path that takes the least time. This is a very profound idea! There is something very deep in it. It also gives the experimentally observed results! Lets apply it to reflection and see what results:

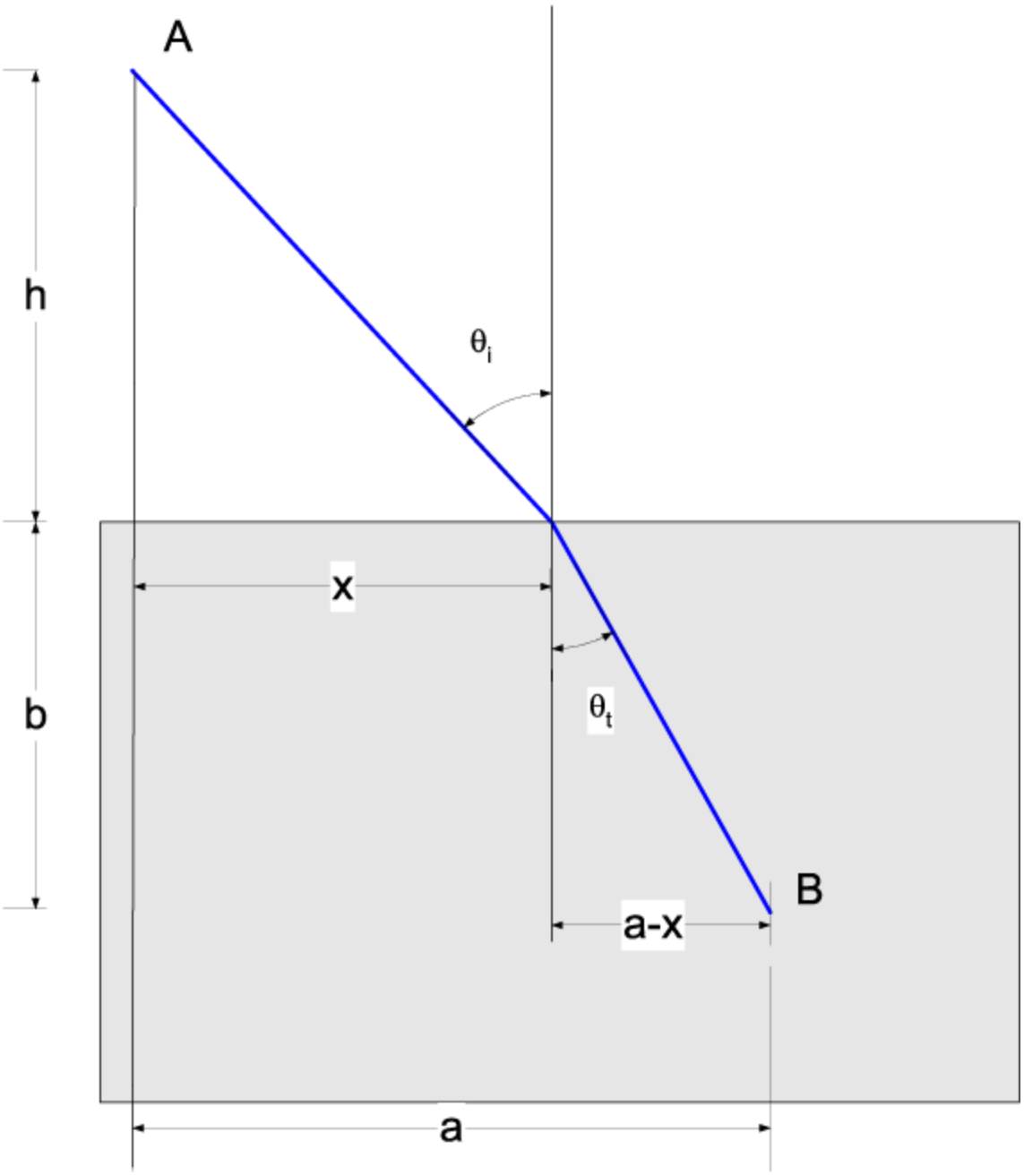


We want to find the length (which is the same as time times the speed) AEB. To do this we construct a fake point B' which is on the other side of the surface the same perpendicular distance from the surface such that the line BB' is a perpendicular to the surface. Then clearly the length AEB

equals the length AEB'. So which point on the surface gives the shortest path to B, the one that gives the shortest path to B' and that clearly lies on the straight line AB'. I have labeled this point C.

Now clearly $\theta_r = \theta'_r$ and also $\theta'_r = \theta_i$ so we get $\theta_r = \theta_i$

Now lets apply Fermat's principle to refraction. Look at the next figure:



We want the shortest time from A to B. Clearly that is

$$t = \frac{\sqrt{h^2 + x^2}}{v_i} + \frac{\sqrt{b^2 + (a-x)^2}}{v_t}$$

To find the minimum we want to solve for x such that $\frac{\partial t}{\partial x} = 0$ Thus

$$\frac{\mathrm{d}t}{\mathrm{d}x} = \frac{x}{v_i\sqrt{h^2+x^2}} + \frac{-(a-x)}{v_t\sqrt{b^2+(a-x)^2}} = 0$$

which is obviously

$$\frac{\sin\theta_i}{v_i} = \frac{\sin\theta_t}{v_t}$$

or Snell's law

$$n_t \sin\theta_t = n_i \sin\theta_i$$

If light travels via many different media then the time is

$$t = \frac{d_1}{v_1} + \frac{d_2}{v_2} + \frac{d_3}{v_3} + \cdots + \frac{d_m}{v_m} +$$

or we can rewrite this as

$$t = \frac{1}{c}\sum_{i=1}^{m} n_i d_i$$

The quantity $\sum_{i=1}^{m} n_i d_i$ is the optical path length (OPL). For a continuously varying medium then the summation becomes (for light traveling from S to P)

$$OPL = \int_S^P n(s)\mathrm{d}s$$

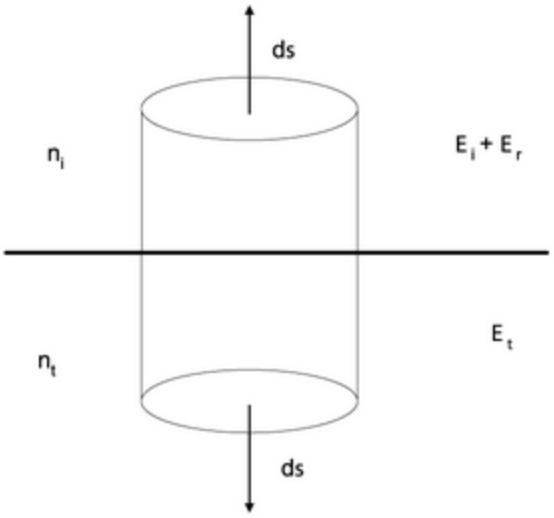
Fermat's principle could be restated that we minimize the OPL In fact this is inadequate, for example one can construct an example where the optical path length is not the minimum.(See for example figure 4.37 in the book "Optics" by Hecht (Fourth Edition).The correct statement of Fermat's principle is that there is a stationary point in the optical path length. (Ie. its derivative is zero).

Electromagnetism at an Interface

EM at an interface

We want to understand with Electromagnetism what happens at a surface. From Maxwell's equations we can understand what happens to the components of the $\vec{E}$ and $\vec{B}$ fields: First lets look at the $\vec{E}$ field using Gauss' law. Recall

$$\oint \varepsilon \vec{E} \cdot d\vec{s} = \int \rho \mathbb{d} V$$



Consider the diagram, the field on the incident side is $\vec{E}_i + \vec{E}_r$. On the transmission side, the field is $\vec{E}_t$. We can collapse the cylinder down so that it is a pancake with an infinitely small height. When we do this there are no field lines through the side of the cylinder. Thus there is only a flux through the top and the bottom of the cylinder and we have;

$$\oint \varepsilon \vec{E} \cdot d\vec{s} = \oint \left[\varepsilon_i \left(E_{i\perp} + E_{r\perp}\right) - \varepsilon_t E_{t\perp}\right] \mathbb{d} s = 0.$$

I have set $\int \rho \mathbb{d}V = 0$ since we will only consider cases without free charges. So we have

$$\varepsilon_i E_{i\perp} + \varepsilon_i E_{r\perp} = \varepsilon_t E_{t\perp}$$

if $\widehat{u}_n$ is a unit vector normal to the surface this can be written

$$\varepsilon_i \widehat{u}_n \cdot \vec{E}_i + \varepsilon_i \widehat{u}_n \cdot \vec{E}_r = \varepsilon_t \widehat{u}_n \cdot \vec{E}_t$$

Similarly Gauss' law of Magnetism

$$\oint \vec{B} \cdot d\vec{s} = 0$$

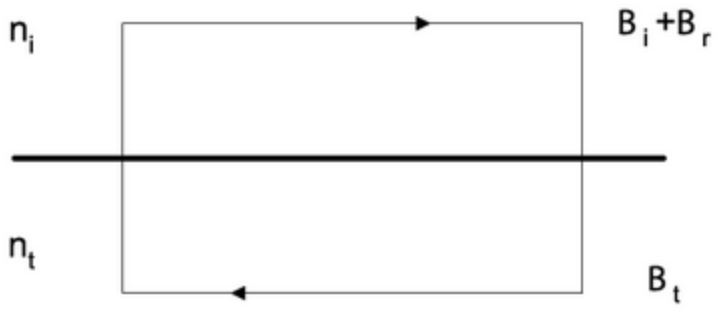
gives

$$B_{i\perp} + B_{r\perp} = B_{t\perp}$$

or

$$\widehat{u}_n \cdot \vec{B}_i + \widehat{u}_n \cdot \vec{B}_r = \widehat{u}_n \cdot \vec{B}_t$$

Amperes law can also be applied to an interface.Then



$$\int \frac{\vec{B}}{\mu} \cdot d\vec{l} = \int \vec{j} \cdot d\vec{s} + \frac{\partial}{\partial t} \int \varepsilon \vec{E} \cdot d\vec{s}$$

(note that in this case $d\vec{s}$ is perpendicular to the page)

Now we will not consider cases with surface currents. Also we can shrink the vertical ends of the loop so that the area of the box is 0 so that $\int \varepsilon \vec{E} \cdot d\vec{s} = 0$. Thus we get at a surface

$$\frac{B_{i\|} + B_{r\|}}{\mu_i} = \frac{B_{t\|}}{\mu_t}$$

or

$$\frac{\hat{u}_n \times \vec{B}_i}{\mu_i} + \frac{\hat{u}_n \times \vec{B}_r}{\mu_r} = \frac{\hat{u}_n \times \vec{B}_t}{\mu_t}$$

Similarly we can use Faraday's law

$$\int \vec{E} \cdot d\vec{l} = -\frac{\partial}{\partial t} \int \vec{B} \cdot d\vec{s}$$

and play the same game with the edges to get

$$E_{i\|} + E_{r\|} = E_{t\|}$$

or

$$\hat{u}_n \times \vec{E}_i + \hat{u}_n \times \vec{E}_r = \hat{u}_n \times \vec{E}_t$$

(notice ε does not appear)

In summary we have derived what happens to the $\vec{E}$ and $\vec{B}$ fields at the interface between two media:

$$\varepsilon_i \hat{u}_n \cdot \vec{E}_i + \varepsilon_i \hat{u}_n \cdot \vec{E}_r = \varepsilon_t \hat{u}_n \cdot \vec{E}_t$$

$$\widehat{u}_n \cdot \vec{B}_i + \widehat{u}_n \cdot \vec{B}_r = \widehat{u}_n \cdot \vec{B}_t$$

$$\frac{\widehat{u}_n \times \vec{B}_i}{\mu_i} + \frac{\widehat{u}_n \times \vec{B}_r}{\mu_i} = \frac{\widehat{u}_n \times \vec{B}_t}{\mu_t}$$

$$\widehat{u}_n \times \vec{E}_i + \widehat{u}_n \times \vec{E}_r = \widehat{u}_n \times \vec{E}_t$$

Snell's Law

Snell's Law

Consider an electromagnetic wave impinging upon an interface:

$$\widehat{u}_n \times \vec{E}_{0i} e^{i(\vec{k}_i \cdot \vec{r} + \omega_i t)} + \widehat{u}_n \times \vec{E}_{0r} e^{i(\vec{k}_r \cdot \vec{r} + \omega_r t + \delta_r)} = \widehat{u}_n \times \vec{E}_{0t} e^{i(\vec{k}_t \cdot \vec{r} + \omega_t t + \delta_t)}$$

Where $(\vec{k_i}, \omega_i)$ describes the incoming wave, $(\vec{k_r}, \omega_r, \delta_r)$ the reflected wave, and $(\vec{k_t}, \omega_t, \delta_t)$ the transmitted wave. At the interface (ie. at points where the vector $\vec{r}$ points to the plane of the interface), all the waves must be in phase with each other. This means that the frequencies must all be equal and there can be no arbitrary phase between the waves. The net result of this is that we must have (for an interface passing through the origin):

$$\vec{k}_i \cdot \vec{r} = \vec{k}_r \cdot \vec{r} = \vec{k}_t \cdot \vec{r}$$

from which we get

$$k_i \sin\theta_i = k_r \sin\theta_r.$$

It is important to note now that we are doing this at the interface. We have chosen a coordinate system so that the interface is at $y = 0$ and contains the origin. This implies that the vector $\vec{r}$ is lying in the plane of the interface at the point where we say that the above is true.

Finally, since the incident and reflected waves are in the same medium we must have $k_i = k_r$ and thus

$$\theta_i = \theta_r$$

Also, we get that $\vec{k}_i, \vec{k}_r, \widehat{u}$ all line in a plane (because $(\vec{k}_i - \vec{k}_r) \cdot \vec{r} = 0$ defines a plane). We also have

$$\widehat{u} \times \left(\vec{k}_i - \vec{k}_t\right) = 0$$

and following the same arguments find that $\vec{k}_i, \vec{k}_r, \vec{k}_t, \widehat{u}$ all line in a plane and that

$$k_i \sin\theta_i = k_t \sin\theta_t.$$

Now we know that $\omega_i = \omega_t$ so we can multiply both sides by c/ω_i and get

$$n_i \sin\theta_i = n_t \sin\theta_t$$

Digression, further justifying the above

At the interface, which we will set to $y = 0$ for convenience (you can always switch back to any coordinate system afterwards. It is good practice to choose the coordinate system that makes your problem easy)

$$\begin{aligned}\left[\vec{k}_i \cdot \vec{r} + \omega_i t\right]\Big|_{y=0} &= \left[\vec{k}_r \cdot \vec{r} + \omega_r t + \delta_r\right]\Big|_{y=0} \\ &= \left[\vec{k}_t \cdot \vec{r} + \omega_t t + \delta_t\right]\Big|_{y=0}\end{aligned}$$

now this must be true for all $\vec{r}$ on the surface and for all t so we must have

$$\omega_i = \omega_r = \omega_t$$

So now we have

$$\begin{aligned}\left[\vec{k}_i \cdot \vec{r} + \omega_i t\right]\Big|_{y=0} &= \left[\vec{k}_r \cdot \vec{r} + \omega_i t + \delta_r\right]\Big|_{y=0} \\ &= \left[\vec{k}_t \cdot \vec{r} + \omega_i t + \delta_t\right]\Big|_{y=0}\end{aligned}$$

which can be written

$$\left[\vec{k}_i \cdot \vec{r}\right]\Big|_{y=0} = \left[\vec{k}_r \cdot \vec{r} + \delta_r\right]\Big|_{y=0} = \left[\vec{k}_t \cdot \vec{r} + \delta_t\right]\Big|_{y=0}$$

So now we can write

$$\left[\left(\vec{k}_i - \vec{k}_r\right) \cdot \vec{r}\right]\Big|_{y=0} = \delta_r$$

Since the interface passes through the origin, one of the allowed values of $\vec{r}$ is 0. So this is only true if $\delta_r = 0$. (If the interface does not include the origin then you can not make this simplification, but clearly we can always choose a coordinate system such that this is true and thereby simplify our lives) likewise we could have written

$$\left[\left(\vec{k}_i - \vec{k}_r\right) \cdot \vec{t}\right]\Big|_{y=0} = \delta_t$$

and applied the same argument to get $\delta_t = 0$. Lets just use this henceforth and thus write:

$$\left[\vec{k}_i \cdot \vec{r}\right]\Big|_{y=0} = \left[\vec{k}_r \cdot \vec{r}\right]\Big|_{y=0} = \left[\vec{k}_t \cdot \vec{r}\right]\Big|_{y=0}$$

The Fresnel Equations

Snell's law, and the law of reflection are very useful for describing what happens to a ray of light at an interface. However we want to acheive a deeper understanding of what is happening to the electric field at the interface. We want to derive the relationships between the $\vec{E}$'s and $\vec{B}$'s. Hence we derive the Fresnel equations.

At an interface Any polarization of the fields can be broken down into simple components (which we will see later). Thus there are just two cases that we have to treat. Both of these cases are drawn at the scienceworld description of the Fresnel Equations http://scienceworld.wolfram.com/physics/FresnelEquations.html. We will first treat the case where the $\vec{E}$ field is perpendicular to the plane of incidence. This corresponds to figure 20.1 in the book "Introduction to Optics" by Pedrotti and Pedrotti (second edition) and to figure 4.39 in the book "Optics" by Hecht (fourth edition). It is the second figure on the scienceworld web page. (It is important to note that while these three sources use similar conventions there are others, for example in this case we could have chosen to have the incoming $\vec{B}$ field pointing upwards in the drawing. This would lead to different signs in the resulting equations.). We have in this case ($\vec{E}$ perpendicular):

$$\widehat{k} \times \vec{E} = \mathrm{v}\vec{B}$$

$$\widehat{k} \cdot \vec{E} = 0$$

Also we know that

$$E_{i\|} + E_{r\|} = E_{t\|}$$

(as was shown earlier) but since in this case the $\vec{E}$ field is parallel to the interface we can write;

$$\vec{E}_{0i} + \vec{E}_{0r} = \vec{E}_{0t}.$$

Also we showed before that

$$\frac{B_{i\|} + B_{r\|}}{\mu_i} = \frac{B_{t\|}}{\mu_t}$$

(where the $\|$ is w.r.t. the surface). Note in the figure how the $\vec{B}$ field changes direction and that the positive x direction is to the right in the figure. It is important to point out here, that we are evaluating things at the interface. At the interface the vector $\vec{r}$ points to the plane of the interface. We have chosen a coordinate system so that the interface that lies at $z = 0$ and contains the origin.

Then we rewrite the equation as

$$\frac{B_i}{\mu_i}\cos\theta_i - \frac{B_r}{\mu_i}\cos\theta_r = \frac{B_t}{\mu_t}\cos\theta_t$$

You may ask, why cosine for a cross product? The cross product uses the complements of θ_i and θ_t which why it ends up being a cosine and not a sine. Using the law of reflection to subsitute for θ_r

$$\frac{B_i}{\mu_i}\cos\theta_i - \frac{B_r}{\mu_i}\cos\theta_i = \frac{B_t}{\mu_t}\cos\theta_t$$

use $B = E/\mathrm{v}$

$$\frac{E_i}{\mathrm{v}_i\mu_i}\cos\theta_i - \frac{E_r}{\mathrm{v}_r\mu_i}\cos\theta_i = \frac{E_t}{\mathrm{v}_t\mu_t}\cos\theta_t$$

use $\mathrm{v}_i = \mathrm{v}_r$

$$\frac{1}{\mu_i\mathrm{v}_i}\left(E_i - E_r\right)\cos\theta_i = \frac{1}{\mu_t\mathrm{v}_t}E_t\cos\theta_t$$

Now at the interface (using the arguments we have used before) we can write

$$\frac{1}{\mu_i \mathbf{v}_i}\left(E_{0i}-E_{0r}\right)\cos\theta_i = \frac{1}{\mu_t \mathbf{v}_t}E_{0t}\cos\theta_t$$

$$\frac{n_i}{\mu_i}\left(E_{0i}-E_{0r}\right)\cos\theta_i = \frac{n_t}{\mu_t}E_{0t}\cos\theta_t$$

now we can substitute in $E_{0i}+E_{0r}=E_{0t}$

$$\frac{n_i}{\mu_i}\left(E_{0i}-E_{0r}\right)\cos\theta_i = \frac{n_t}{\mu_t}\left(E_{0i}+E_{0r}\right)\cos\theta_t$$

Then rearrange or

$$\left(\frac{E_{0r}}{E_{0i}}\right)_\perp = \frac{\frac{n_i}{\mu_i}\cos\theta_i - \frac{n_t}{\mu_t}\cos\theta_t}{\frac{n_i}{\mu_i}\cos\theta_i + \frac{n_t}{\mu_t}\cos\theta_t}$$

For the transmission we again start with

$$\frac{n_i}{\mu_i}\left(E_{0i}-E_{0r}\right)\cos\theta_i = \frac{n_t}{\mu_t}E_{0t}\cos\theta_t$$

and eliminate E_{0r} using $E_{0i}+E_{0r}=E_{0t}$ or $E_{0r}=E_{0t}-E_{0i}$

$$\frac{n_i}{\mu_i}\left(2E_{0i}-E_{0t}\right)\cos\theta_i = \frac{n_t}{\mu_t}E_{0t}\cos\theta_t$$

$$2\frac{n_i}{\mu_i}E_{0i}\cos\theta_i = \left(\frac{n_i}{\mu_i}\cos\theta_i + \frac{n_t}{\mu_t}\cos\theta_t\right)E_{0t}$$

$$\left(\frac{E_{0t}}{E_{0i}}\right)_\perp = \frac{2\frac{n_i}{\mu_i}\cos\theta_i}{\frac{n_i}{\mu_i}\cos\theta_i + \frac{n_t}{\mu_t}\cos\theta_t}$$

Now usually we only consider materials where $\mu_i \approx \mu_t \approx \mu_0$ so these equations simplify to:

$$r_\perp \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\perp = \frac{n_i \cos\theta_i - n_t \cos\theta_t}{n_i \cos\theta_i + n_t \cos\theta_t}$$

$$t_\perp \equiv \left(\frac{E_{0t}}{E_{0i}}\right)_\perp = \frac{2n_i \cos\theta_i}{n_i \cos\theta_i + n_t \cos\theta_t}$$

If the $\vec{E}$ field is in the plane of incidence (the first figure on the scienceworld web page, figure 4.40 in Hecht, figure 20.2 in Pedrotti and Pedrotti) then the same sort of procedure can be followed (Pedrotti and Pedrott pages 409,410). It is left as a homework problem to show that in this case:

$$r_\parallel \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\parallel = \frac{n_t \cos\theta_i - n_i \cos\theta_t}{n_t \cos\theta_i + n_i \cos\theta_t}$$

$$t_\parallel \equiv \left(\frac{E_{0t}}{E_{0i}}\right)_\parallel = \frac{2n_i \cos\theta_i}{n_t \cos\theta_i + n_i \cos\theta_t}$$

Module 34.xhtml

The Fresnel Equations

$$r_\perp \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\perp = \frac{n_i \cos\theta_i - n_t \cos\theta_t}{n_i \cos\theta_i + n_t \cos\theta_t}$$

$$t_\perp \equiv \left(\frac{E_{0t}}{E_{0i}}\right)_\perp = \frac{2n_i \cos\theta_i}{n_i \cos\theta_i + n_t \cos\theta_t}$$

$$r_\parallel \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\parallel = \frac{n_t \cos\theta_i - n_i \cos\theta_t}{n_t \cos\theta_i + n_i \cos\theta_t}$$

$$t_\parallel \equiv \left(\frac{E_{0t}}{E_{0i}}\right)_\parallel = \frac{2n_i \cos\theta_i}{n_t \cos\theta_i + n_i \cos\theta_t}$$

Why Polaroid sunglasses work

Lets put these equations to work and figure out something practical. Consider light reflecting off a surface, such as the road in front of you when you are driving a car. The light hitting the road surface can have any polarization but that will be some addition of light that has $\vec{E}$ $\perp$ and $\parallel$ to the plane of incidence. From

$$r_\parallel \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\parallel = \frac{n_t \cos\theta_i - n_i \cos\theta_t}{n_t \cos\theta_i + n_i \cos\theta_t}$$

and

$$r_\perp \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\perp = \frac{n_i \cos\theta_i - n_t \cos\theta_t}{n_i \cos\theta_i + n_t \cos\theta_t}$$

we see there is an angle where there is no $\parallel$ and only $\perp$ light, namely

$$n_t \cos\theta_i - n_i \cos\theta_t = 0$$

or

$$n_t \cos\theta_i = n_i \cos\theta_t.$$

Now we can use Snell's law rewritten as

$$\frac{\sin\theta_i}{n_t} = \frac{\sin\theta_t}{n_i}$$

and multiply both sides to get

$$n_t \cos\theta_i = n_i \cos\theta_t$$

$$n_t \cos\theta_i \frac{\sin\theta_i}{n_t} = n_i \cos\theta_t \frac{\sin\theta_t}{n_i}$$

$$\cos\theta_i \sin\theta_i = \cos\theta_t \sin\theta_t$$

which can only be true if

$$\theta_t = \pi/2 - \theta_i.$$

In this case Snell's law can be written

$$\begin{aligned} \frac{n_t}{n_i} &= \frac{\sin\theta_i}{\sin\theta_t} \\ &= \frac{\sin\theta_i}{\sin(\pi/2 - \theta_i)} \\ &= \tan\theta_i \end{aligned}$$

The angle that gives this effect is known as Brewster's angle

$$\theta_i \equiv \theta_{Brewster} = \tan^{-1}\frac{n_t}{n_i}$$

At this angle light is completely polarized, it only has $\vec{E} \perp$ to the plane of incidence (or parallel to the surface). Thus the glare you get from reflected light tends to be polarized in this way. In 1929 Edwin Land invented a method for making celluoid filters to filter out light with given

polarizations. He then manufactured sunglasses with these polarizers lined up to filter out the $E_\perp$ light and thereby reduce glare.

Other cases

Look at the Fresnel equations again and examine what happens when θ_i approaches 90 degrees. The reflection approaches 1 (ignore the signs) Thus at Glancing incidence you get lots of reflection. In fact X-ray telescopes use this to focus the x-rays onto their detector.

Another effect, if $n_i > n_t$ then there is an angle of incidence beyond which light is only reflected. That is the angle where $\theta_t = 90$ degrees.

$$n_t \sin\theta_t = n_i \sin\theta_i$$

$$\theta_t \to \pi/2$$

$$\sin\theta_t \to 1$$

$$\sin\theta_i \to \frac{n_t}{n_i}$$

The critical angle at which this occurs is

$$\theta_c = \sin^{-1}\frac{n_t}{n_i}$$

Evanescent Wave

We saw that when the index of refraction of the incident material is greater than the transmitting material we can get total internal reflection at the critical angle. An interesting question is "what happens at larger angles of incidence?" This actually is somewhat subtle. From simple trigonometry we know that

$$\cos\theta_t = \sqrt{1 - \sin^2\theta_t}.$$

We also know from Snell's law that

$$\sin\theta_t = \frac{n_i}{n_t}\sin\theta_i$$

so we have

$$\cos\theta_t = \sqrt{1 - \frac{n_i^2}{n_t^2}\sin^2\theta_i}.$$

So we see that if $n_i > n_t$ $\cos\theta_t$ can become an imaginary number! For convenience we will write this as

$$\cos\theta_t = i\sqrt{\frac{n_i^2}{n_t^2}\sin^2\theta_i - 1}$$

Now lets write down the expression for the transmitted wave:

$$E_t = E_{0t}e^{i(\vec{K}_t\cdot\vec{r}-\omega t)}$$

For simplicity we will assume that the interface lies in the $y = 0$ plane and thus the y direction is normal to the interface. Also, we assume the $z = 0$ plane defines then plane of incidence. Then we can write

$$\vec{K}_t = (K_{tx}, K_{ty}, 0)$$

or

$$\vec{K}_t = (K_t \sin\theta_t, K_t \cos\theta_t, 0).$$

Also

$$\vec{r} = (x, y, 0).$$

So now we can write that the wave as

$$E_t = E_{0t} e^{i(\vec{K}_t \cdot \vec{r} - \omega t)}$$

$$E_t = E_{0t} e^{iK_t \sin\theta_t x} e^{iK_t \cos\theta_t y} e^{-i\omega t}$$

or

$$E_t = E_{0t} e^{iK_t \sin\theta_t x} e^{-\sqrt{\frac{n_i^2}{n_t^2}\sin^2\theta_i - 1}\, y} e^{-i\omega t}.$$

It is interesting to note the effect of the term

$$e^{-\sqrt{\frac{n_i^2}{n_t^2}\sin^2\theta_i - 1}\, y}$$

in that expression. This is an exponential decay. The amplitude of the wave drops rapidly to zero.

So there is a transmitted wave but its amplitude drops precipitously. This is referred to as the evanescent wave.

Phase Changes

$$r_\perp \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\perp = \frac{n_i \cos\theta_i - n_t \cos\theta_t}{n_i \cos\theta_i + n_t \cos\theta_t}$$

$$t_\perp \equiv \left(\frac{E_{0t}}{E_{0i}}\right)_\perp = \frac{2n_i \cos\theta_i}{n_i \cos\theta_i + n_t \cos\theta_t}$$

$$r_\parallel \equiv \left(\frac{E_{0r}}{E_{0i}}\right)_\parallel = \frac{n_t \cos\theta_i - n_i \cos\theta_t}{n_t \cos\theta_i + n_i \cos\theta_t}$$

$$t_\parallel \equiv \left(\frac{E_{0t}}{E_{0i}}\right)_\parallel = \frac{2n_i \cos\theta_i}{n_t \cos\theta_i + n_i \cos\theta_t}$$

We can rewrite these equations using Snell's Law to eliminate the $\cos\theta_t$ term. From simple trigonometry we know that

$$\cos\theta_t = \sqrt{1 - \sin^2\theta_t}.$$

We also know from Snell's law that

$$\sin\theta_t = \frac{n_i}{n_t}\sin\theta_i$$

so we have

$$\cos\theta_t = \sqrt{1 - \frac{n_i^2}{n_t^2}\sin^2\theta_i}.$$

We can substitute this into

$$
\begin{aligned}
r_\perp &= \frac{n_i\cos\theta_i - n_t\sqrt{1-\frac{n_i^2}{n_t^2}\sin^2\theta_i}}{n_i\cos\theta_i + n_t\sqrt{1-\frac{n_i^2}{n_t^2}\sin^2\theta_i}} \\
&= \frac{n_i\cos\theta_i - \sqrt{n_t^2 - n_i^2\sin^2\theta_i}}{n_i\cos\theta_i + \sqrt{n_t^2 - n_i^2\sin^2\theta_i}} \\
&= \frac{\cos\theta_i - \sqrt{\frac{n_t^2}{n_i^2} - \sin^2\theta_i}}{\cos\theta_i + \sqrt{\frac{n_t^2}{n_i^2} - \sin^2\theta_i}}
\end{aligned}
$$

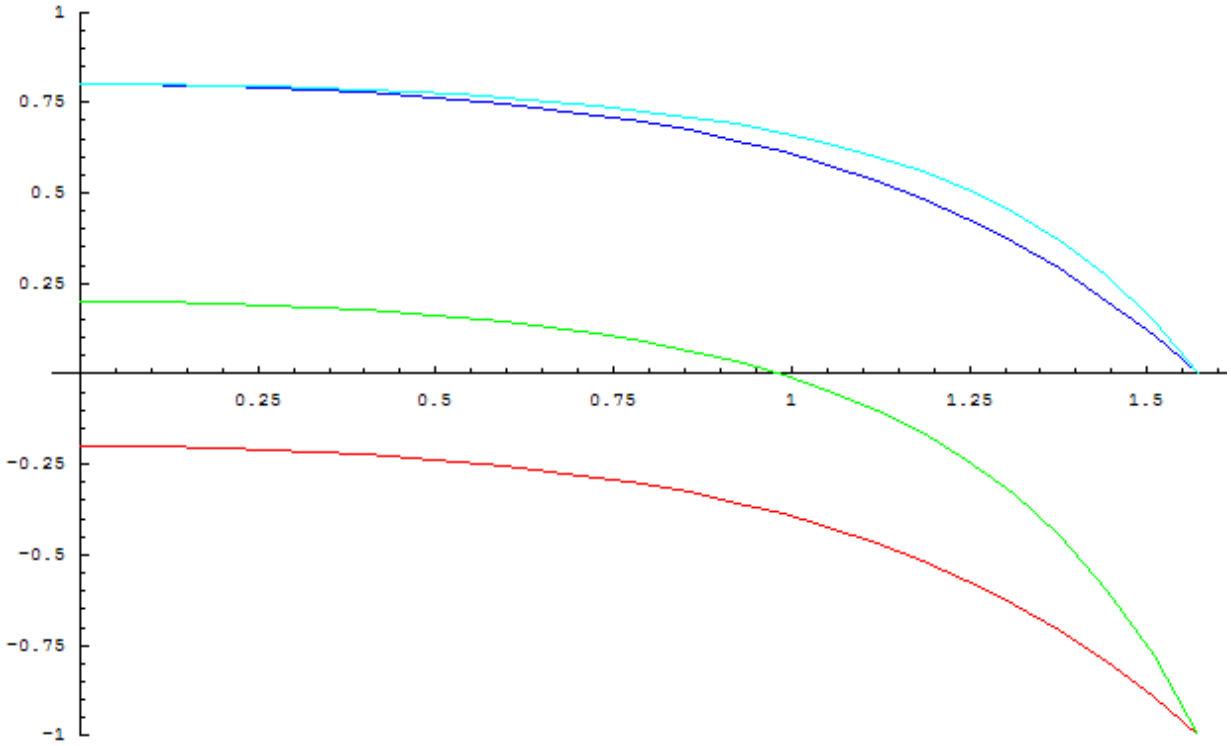
Similarly we can derive that

$$
r_\parallel = \frac{\frac{n_t^2}{n_i^2}\cos\theta_i - \sqrt{\frac{n_t^2}{n_i^2} - \sin^2\theta_i}}{\frac{n_t^2}{n_i^2}\cos\theta_i + \sqrt{\frac{n_t^2}{n_i^2} - \sin^2\theta_i}}
$$

$$
t_\perp = \frac{2\cos\theta_i}{\cos\theta_i + \sqrt{\frac{n_t^2}{n_i^2} - \sin^2\theta_i}}
$$

$$
t_\parallel = \frac{2\frac{n_t}{n_i}\cos\theta_i}{\frac{n_t^2}{n_i^2}\cos\theta_i + \sqrt{\frac{n_t^2}{n_i^2} - \sin^2\theta_i}}
$$

This form allows us to easily plot the coefficients for different values of θ_i. For example,here are the coefficients for the case where $\frac{n_t}{n_i} = 1.5$.



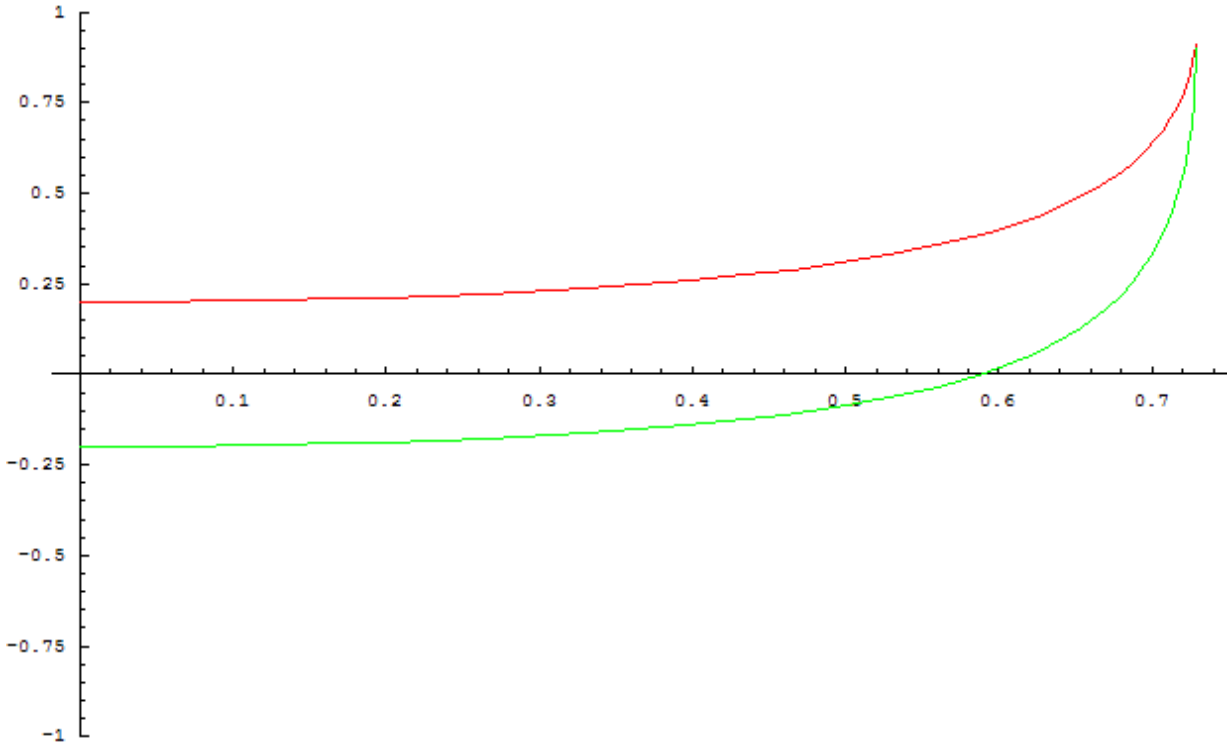
The transmmission and reflection coefficients for the case where the ratio of transmitted to incident indices of refraction is 1.5. The top two curves are transmission. The lower two are reflection, with red being for the E field transverse to the plane of incidence.

It is interesting to note that the sign of the coefficient can change on reflection.

$$E_r = -|r|E = e^{i\pi}|r|E_0 e^{i(\vec{K}\cdot\vec{r}-\omega t)} = |r|E_0 e^{i(\vec{K}\cdot\vec{r}-\omega t+\pi)}$$

This corresponds to a phase change by π upon reflection.

We can also look at what happens to the reflection coefficients when $\frac{n_i}{n_t} = 1.5$.



The reflection coefficients for the case where the ratio of the incident to the transmitted incidence of reflection is 1.5.

That is going from a high index of refraction material to a lesser. In this case we see at the critical angle we get total internal reflection. What happens to the phase here is complicated.

When we have $\frac{n_i}{n_t} > 1$ (or $\frac{n_t}{n_i} < 1$) it is convenient to write

$$r_\perp = \frac{\cos\theta_i - i\sqrt{\sin^2\theta_i - \frac{n_t^2}{n_i^2}}}{\cos\theta_i + i\sqrt{\sin^2\theta_i - \frac{n_t^2}{n_i^2}}}$$

Now to understand what this implies we need to digress a little. Recall that

$$e^{i\alpha} = \cos\alpha + i\sin\alpha.$$

We could have written this as

$$e^{i\alpha} = a + ib$$

then we see that

$$\alpha = \tan^{-1} \frac{b}{a}.$$

Now consider

$$\frac{\cos\alpha - i\sin\alpha}{\cos\alpha + i\sin\alpha} = \frac{e^{-i\alpha}}{e^{i\alpha}} = e^{-2i\alpha}$$

Now looking back at $r_\perp$ we see that

$$r_\perp = e^{-i\varphi}$$

where

$$\tan\frac{\varphi}{2} = \frac{\sqrt{\sin^2\theta_i - \frac{n_t^2}{n_i^2}}}{\cos\theta_i}.$$

We could go through a similar excersize for $r_\parallel$and get the same result with

$$\tan\frac{\varphi}{2} = \frac{\sqrt{\sin^2\theta_i - \frac{n_t^2}{n_i^2}}}{\frac{n_t^2}{n_i^2}\cos\theta_i}.$$

Figure 20-8 in Pedrotti and Pedrotti summarizes all the possible phase changes.

Module 35.xhtml

Reflectance and Transmittance

Lets remember what Irradiance is: We have the Poynting vector

$$\vec{S} = c^2 \varepsilon_0 \vec{E}_0 \times \vec{B}_0 e^{i2\left(\vec{k}\cdot\vec{r}-\omega t\right)}$$

But the problem is that this varies rapidly in time. So we define the Irradiance

$$I = \langle S \rangle_T$$

which has units Watts per meter squared and can also be called the radiant flux density. We showed in lecture that this is (where I drop the T because it is tiresome to write)

$$< S >= c^2 \varepsilon_0 E_0 B_0 / 2$$

which can also be written

$$< S >= c \varepsilon_0 E_0^2 / 2$$

We need to make one minor modification though, the above presumes we are in free space. So we modify the irradiance to take into account that different media have different speeds of light. (Now we write $\mathbf{v}$ instead of c because we are not assuming free space)

$$I =< S >= \frac{\mathbf{v}\varepsilon}{2} E_0^2$$

The reflectance is the ratio of the reflected power to the incident power is

$$R = \frac{I_r A \cos\theta_r}{I_i A \cos\theta_i}$$

A is the area illuminated by the electomagnetic radiation and θ_r, θ_i are the reflected and incident angles. But we know that $\theta_r = \theta_i$ so

$$R = \frac{\frac{\mathrm{v}_r \varepsilon_r}{2} E_{0r}^2}{\frac{\mathrm{v}_i \varepsilon_i}{2} E_{0i}^2}$$

and the media are the same so

$$R = \left(\frac{E_{0r}}{E_{0i}} \right)^2 = r^2.$$

Likewise the transmittance (using $\mu_0 \approx \mu_t \approx \mu_i$) is

$$\begin{aligned} T &= \frac{I_t A \cos\theta_t}{I_i A \cos\theta_i} \\ &= \frac{I_t \cos\theta_t}{I_i \cos\theta_i} \\ &= \frac{\frac{\mathrm{v}_t \varepsilon_t}{2} E_{0t}^2 \cos\theta_t}{\frac{\mathrm{v}_i \varepsilon_i}{2} E_{0i}^2 \cos\theta_i} \\ &= \frac{\mu_0 \mathrm{v}_t \varepsilon_t E_{0t}^2 \cos\theta_t}{\mu_0 \mathrm{v}_i \varepsilon_i E_{0i}^2 \cos\theta_i}. \end{aligned}$$

Now note

$$\begin{aligned} \mu_0 \varepsilon \mathrm{v} &= \mu \varepsilon \mathrm{v} \\ &= \mu \varepsilon \frac{1}{\sqrt{\mu\varepsilon}} \frac{c}{c} \\ &= \frac{\sqrt{\mu\varepsilon}}{\sqrt{\mu_0 \varepsilon_0}} \frac{1}{c} \\ &= \frac{n}{c}. \end{aligned}$$

So now we can write

$$\begin{aligned} T &= \frac{\mu_0 \mathrm{v}_t \varepsilon_t E_{0t}^2 \cos\theta_t}{\mu_0 \mathrm{v}_i \varepsilon_i E_{0i}^2 \cos\theta_i} \\ &= \frac{n_t E_{0t}^2 \cos\theta_t}{n_i E_{0i}^2 \cos\theta_i} \\ &= \frac{n_t \cos\theta_t}{n_i \cos\theta_i} \left(\frac{E_{0t}}{E_{0i}} \right)^2 \\ &= \frac{n_t \cos\theta_t}{n_i \cos\theta_i} t^2. \end{aligned}$$

This is a more complicated expression than R because1)The speed of energy transmission is affected by the medium2) $\theta_i \neq \theta_t$ so the projected areas normal to the propagation direction are different.

Polarization

Polarization

Recall that we can add waves so lets take a plane wave traveling in the z direction and break it into components.

$$\vec{E}_x = E_{0x}\cos\left(kz - \omega t\right)\hat{\imath}$$

$$\vec{E}_y = E_{0y}\cos\left(kz - \omega t + \delta\right)\hat{}$$

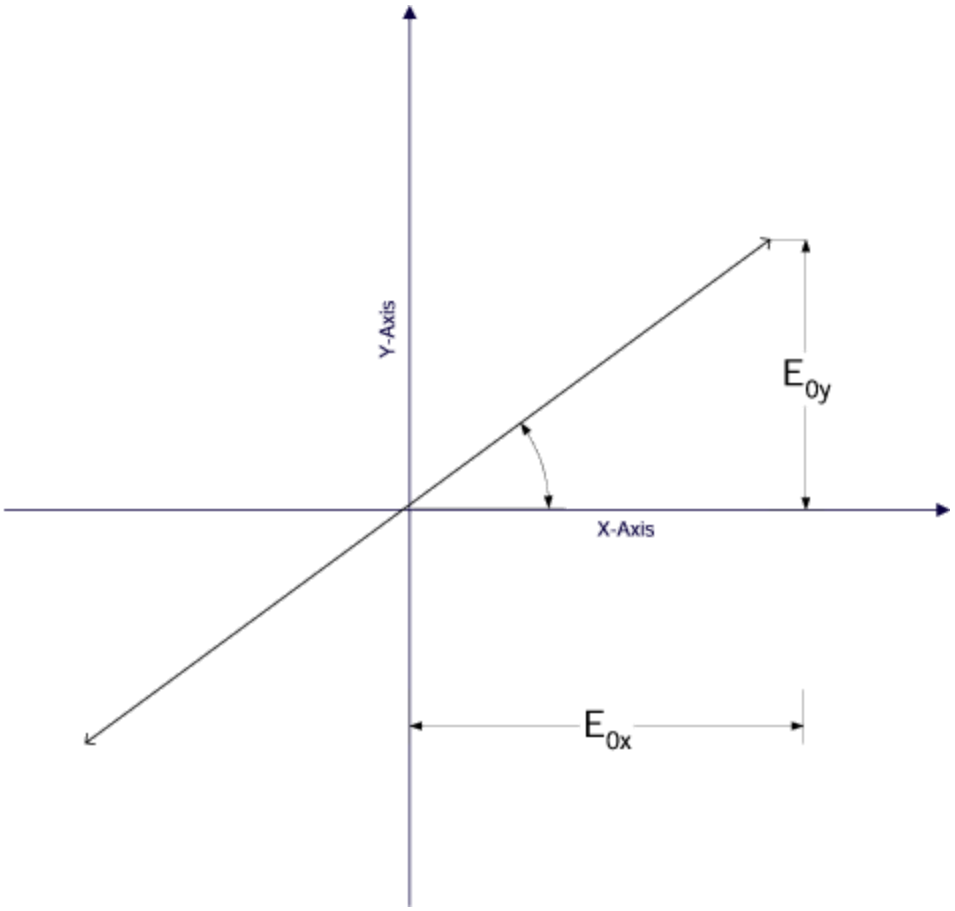
Where δ is some arbitrary phase between the two components. The electric field for the wave is

$$\vec{E} = \vec{E}_x + \vec{E}_y$$

Now there are a number of different cases that arise.

If $\delta = 0,\ 2\pi,\ 4\pi, \ldots$ Then we can write the field as

$$\vec{E} = \left(E_{0x}\hat{\imath} + E_{0y}\hat{}\right)\cos\left(kz - \omega t\right)$$

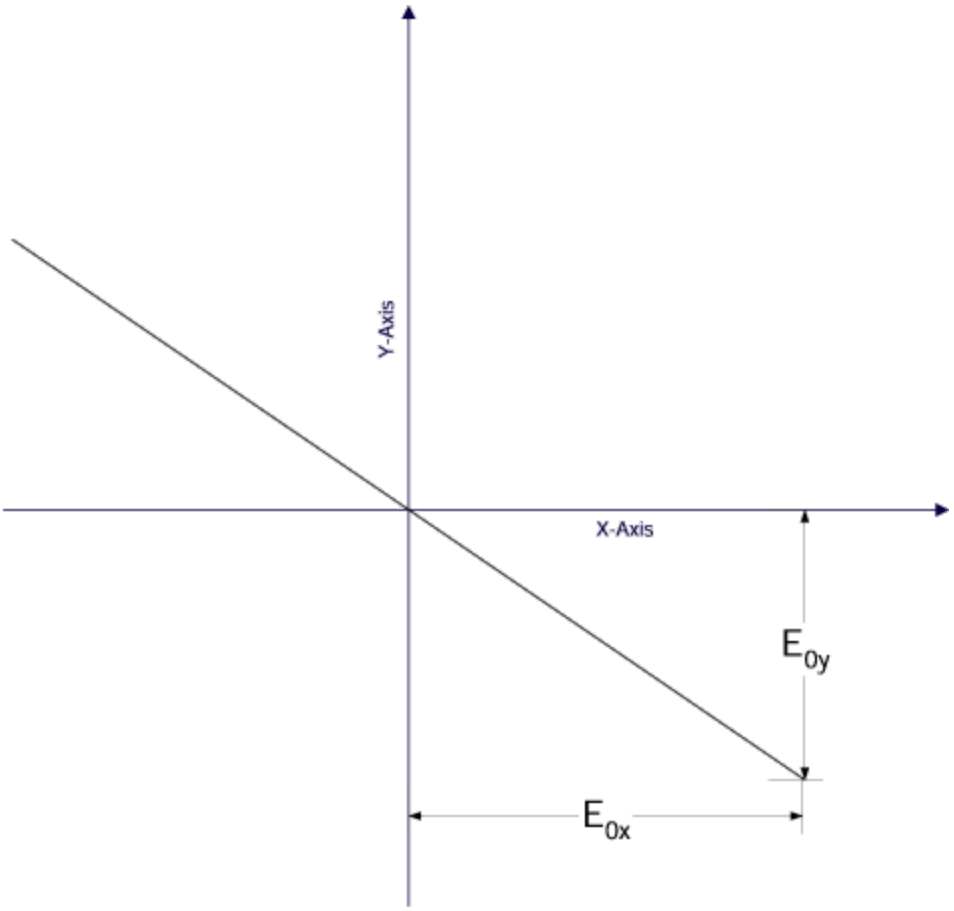


Polarization of light for the case $\vec{E} = (E_{0x}\hat{\imath} + E_{0y}\hat{\ }) \cos(kz - \omega t)$

The field is linearly polarized, that is the E field lies along a straight line.

Likewise If $\delta = \pi,\ 3\pi,\ 5\pi, \ldots$ Then again it is linearly polarized, but is now "flipped"

$$\vec{E} = (E_{0x}\hat{i} - E_{0y}\hat{j}) \cos(kz - \omega t)$$



Polarization of light for the case $\vec{E} = (E_{0x}\hat{\imath} - E_{0y}\hat{\ }) \cos(kz - \omega t)$

Circularly polarized light is a particularly interest example. Let $\delta = -\pi/2$ and $E_{0x} = E_{0y} = E_0$

Then

$$\vec{E}_x = E_0 \cos(kz - \omega t)\hat{\imath}$$

$$\vec{E}_y = E_0 \cos(kz - \omega t - \pi/2)\hat{\ }$$

$$\vec{E}_y = E_0 \sin(kz - \omega t)\hat{\ }$$

or

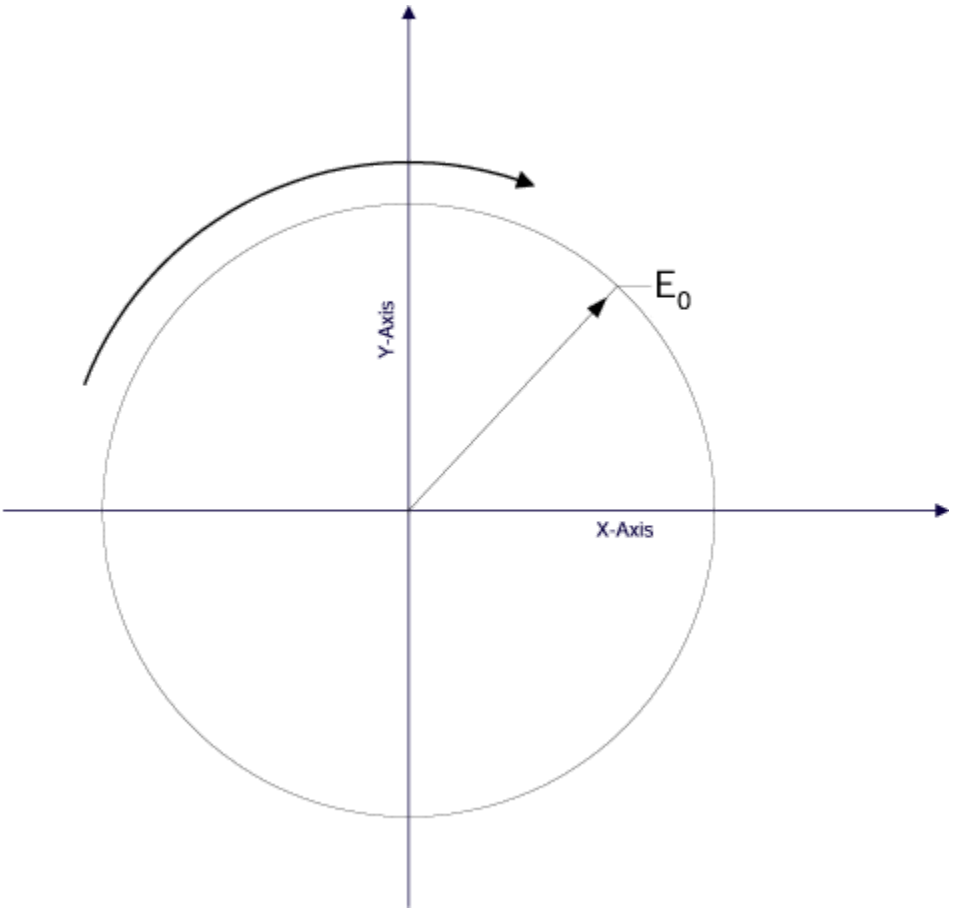
$$\vec{E} = E_0 \left[\cos\left(kz - \omega t\right)\hat{\imath} + \sin\left(kz - \omega t\right)\hat{} \right]$$

The direction of $\vec{E}$ is changing with time. For example consider the case at $z = 0$. Then

$$\vec{E} = E_0 \left[\cos\left(-\omega t\right)\hat{\imath} + \sin\left(-\omega t\right)\hat{} \right]$$

$$\vec{E} = E_0 \left[\cos\left(\omega t\right)\hat{\imath} - \sin\left(\omega t\right)\hat{} \right]$$

In this case the electric field undergoes uniform circular rotation. For light coming out of the page, it will have the motion shown in the figure



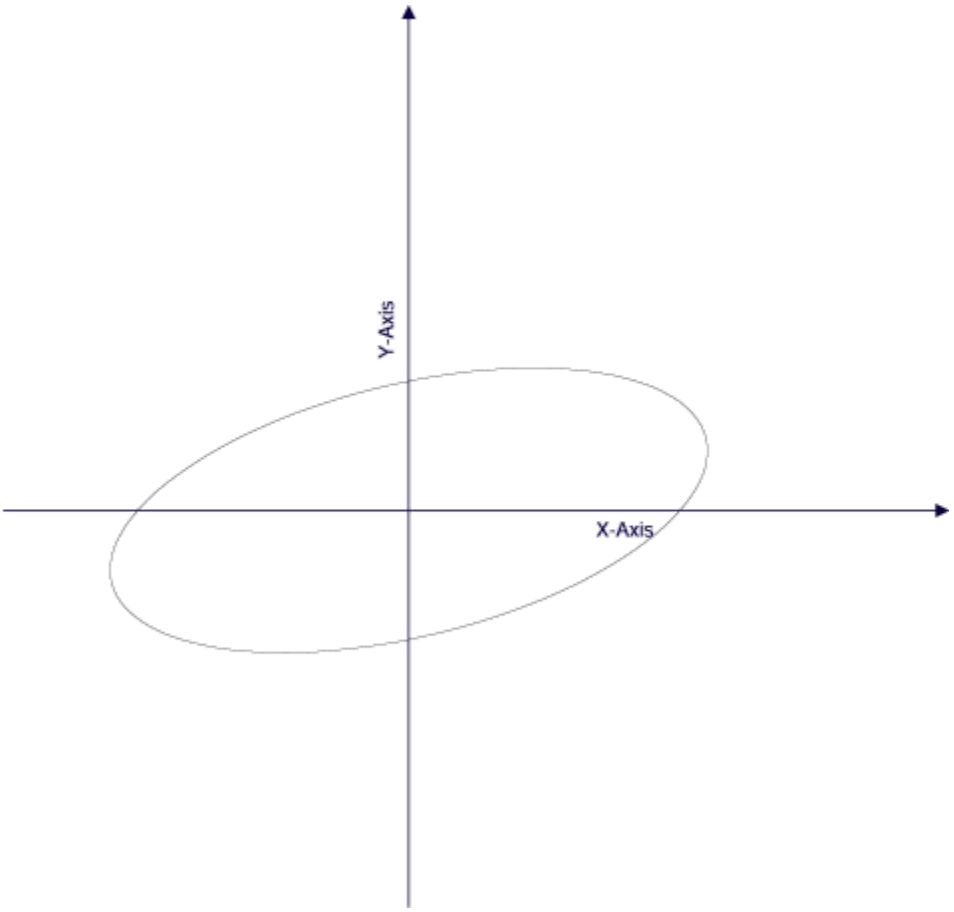
Polarization of light for the case $\vec{E} = E_0 \left[\cos(\omega t)\hat{\imath} - \sin(\omega t)\hat{\jmath}\right]$. I, and most physicists would call the Left Hand Circular polarization (LHC). The thumb points in the direction of the light ray and the fingers curve in the direction of rotation. (This is known as the Angular momentum convention, optical scientists will use the "Optical" convention which is opposite - we will stick to the angular momentum convention.)

Suppose $\delta = \pi/2$ then you get

$$\vec{E} = E_0 \left[\cos(kz - \omega t)\hat{\imath} - \sin(kz - \omega t)\hat{\jmath}\right]$$

This now has the opposite rotation, it is Right Handed.

The most general case of polarization has δ arbitrary and $E_{0x} \neq E_{0y}$ and is elliptically polarized.



Elliptical polarization

Natural light

Natural light is emitted with the $\vec{E}$ field in a mixture of random directions. This in unpolarized light. At any particular instant $\vec{E}$ has a particular direction, but that direction changes rapidly and randomly.

Malus' Law

A polarizer is a device that takes incident natural light and transmits polarized light. For example a linear polarizer will take incident light and select only that component of the light that has its $\vec{E}$ field lined up along the

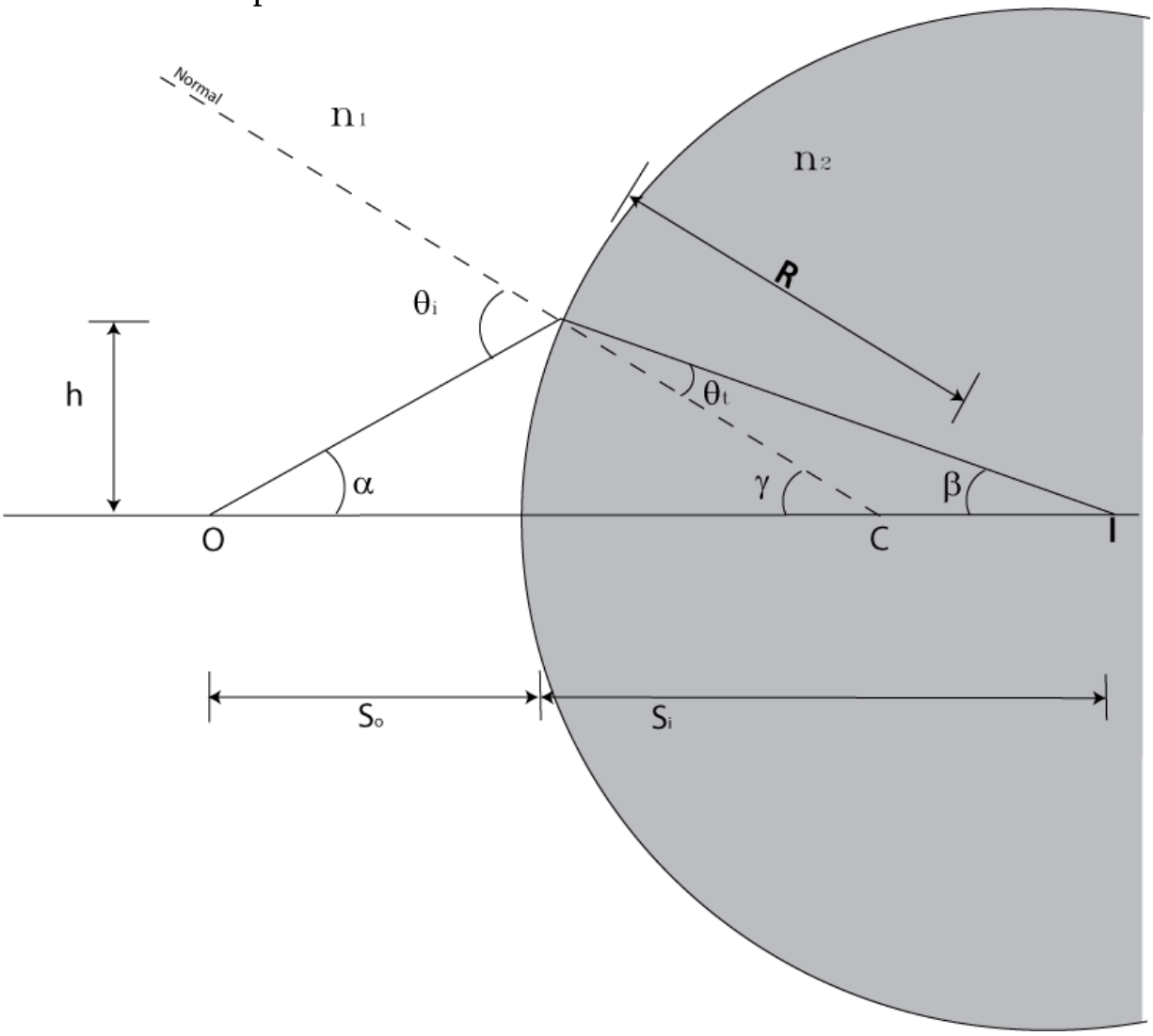
transmission axis. Suppose there is a linear polarizer that transmits light along a particular axis. This is followed by a second linear polarizer that has its transmission axis at a different angle with θ being the angle between the transmission axes. Since $I \sim E^2$ then at the second polarizer

$$I(\theta) = I(0)\cos^2\theta$$

where $I(0)$ is the Irradiance of the light hitting the second polarizer.

Lets consider unpolarized light hitting a linear polarizer. Then half the Irradiance will get through. If this is followed by a second polarizer at 90^0 then no light will pass through the second polarizer. Now what happens if a third polarizer is placed between them?

Refraction at a Spherical Interface



Refraction at a spherical interface. Click on image for larger version.

Look at the figure showing refraction at a sphere.In this figure:

- C is the center of curvature of the spherical surface
- R is the radius of curvature
- O is the position of the Object
- I is the position of the Image
- S_o is the distance of the object from the surface along the optical axis

- S_i is the distance from the surface to the Image
- $n_1 < n2$

There is a ray that strikes the surface at height h. In general rays hitting the surface at different points will be bent to different points along the optical axis. However for small angles we will show they all converge at the same point.So lets use the small angle approximation

$$\tan\alpha = \frac{h}{s_o} \approx \alpha$$

$$\tan\beta = \frac{h}{s_i} \approx \beta$$

$$\tan\gamma = \frac{h}{R} \approx \gamma$$

Now from trigonometry we can see that:

$$\theta_i = \alpha + \gamma$$

$$\gamma = \theta_t + \beta$$

or

$$\theta_t = \gamma - \beta$$

now Snell's law says

$$n_1 \sin\theta_i = n_2 \sin\theta_t$$

or

$$n_1\theta_i \approx n_2\theta_t$$

$$n_1(\alpha + \gamma) = n_2(\gamma - \beta)$$

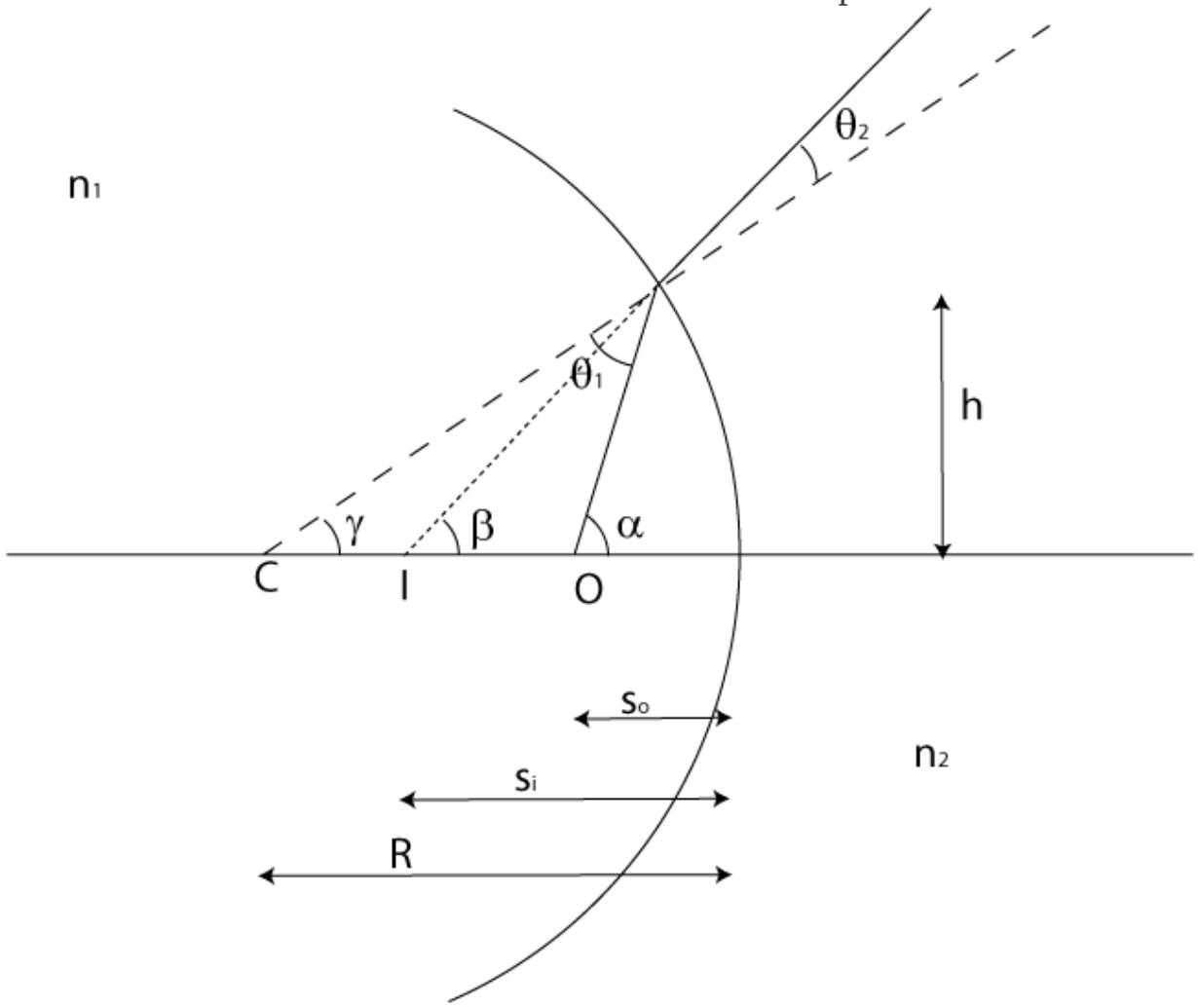
$$\gamma(n_2 - n_1) = n_2\beta + n_1\alpha$$

$$\frac{h}{R}(n_2 - n_1) = n_2\frac{h}{s_i} + n_1\frac{h}{s_o}$$

Now all the h's cancel so there is no dependence on point on surface. That is:

$$\frac{n_2 - n_1}{R} = \frac{n_2}{s_i} + \frac{n_1}{s_o}$$

Now lets consider the case of a concave surface. The picture is



Click to get larger image.

Again we use the small angle approximation and thus we have

$$n_1\theta_1 = n_2\theta_2$$

In this case we also see that

$$\alpha = \theta_1 + \gamma$$

and

$$\beta = \theta_2 + \gamma$$

so we can write

$$n_1(\alpha - \gamma) = n_2(\beta - \gamma)$$

or

$$n_1\left(\frac{h}{s_o} - \frac{h}{R}\right) = n_2\left(\frac{h}{s_i} - \frac{h}{R}\right)$$

or

$$\frac{n_1}{s_o} - \frac{n_2}{s_i} = \left(\frac{n_1 - n_2}{R}\right)$$

However we can make the equation identical to the previous one if we adopt the following sign convention:

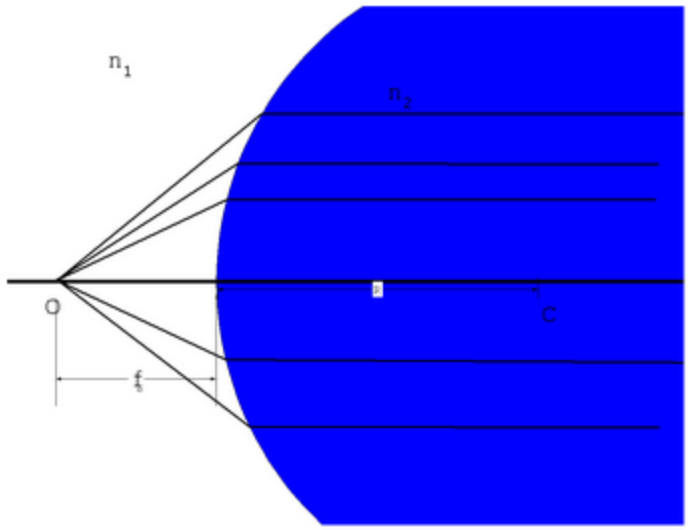
- s_o is positive to the left of the interface
- s_i is positive to the right of the interface
- R is positive when the center of the sphere is to the right of the interface

Then the equation becomes as before

$$\frac{n_2 - n_1}{R} = \frac{n_2}{s_i} + \frac{n_1}{s_o}$$

In this case note that the image is imaginary (whereas in the first case it was real). Note that the actual rays pass through a real image.

The focal point is the object point which causes the image to occur at infinity.



That is all the rays end up traveling parallel to each other. In this case s_i goes to∞ so

$$\frac{n_1}{s_o} + \frac{n_2}{s_i} = \frac{n_2 - n_1}{R}$$

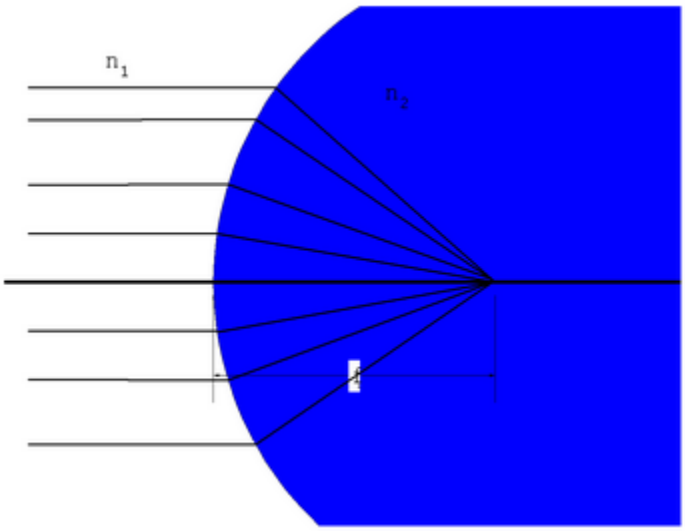
becomes

$$\frac{n_1}{f_o} = \frac{n_2 - n_1}{R}$$

or

$$f_o = \frac{n_1 R}{n_2 - n_1}.$$

Now we can find a focal point to the right of the of the surface by considering parallel rays coming in from the left.



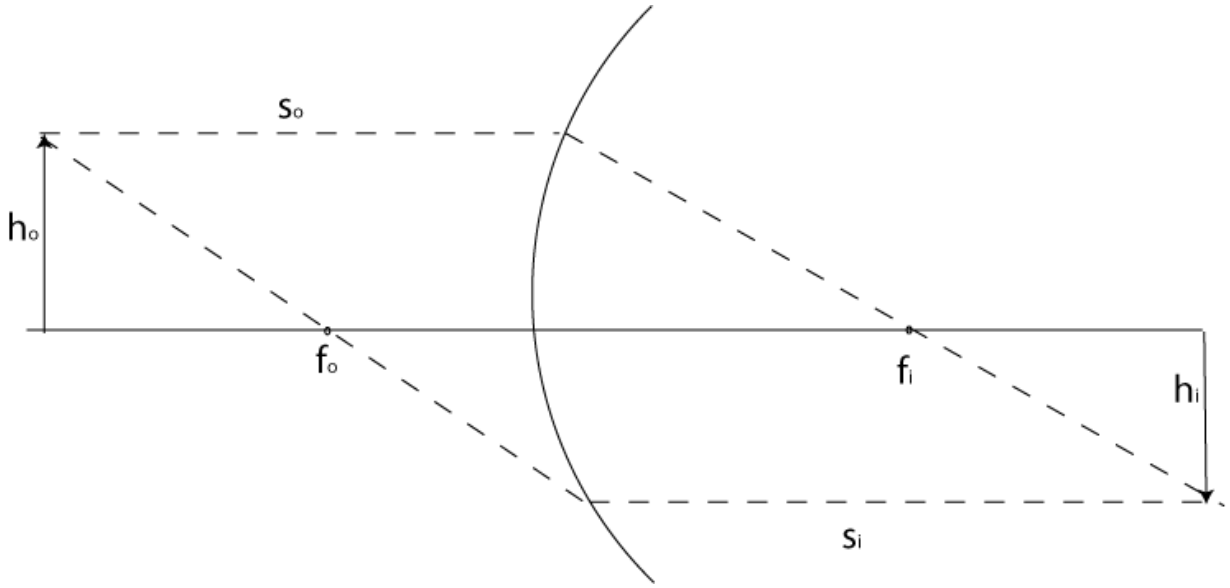
Then we get

$$f_i = \frac{n_2 R}{n_2 - n_1}$$

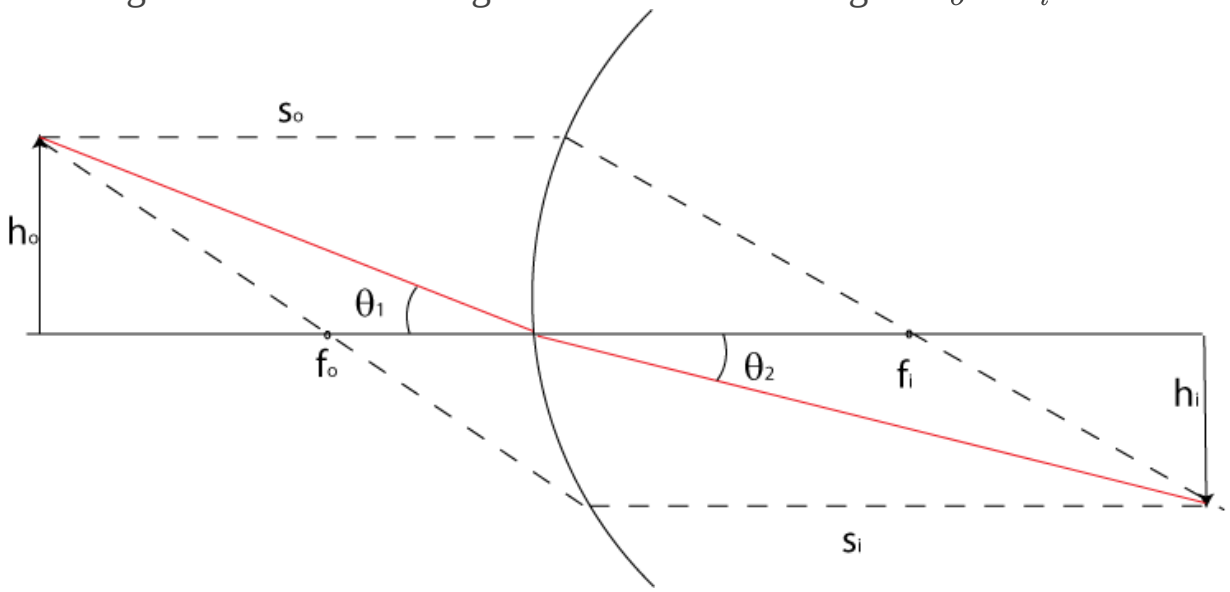
But we do have to expand our sign conventionfor light incident from the left

- s_o is positive to the left of the interface
- s_i is positive to the right of the interface
- R is positive when the center of the sphere is to the right of the interface
- f_o is positive to the left of the interface
- f_i is positive to the right of the interface

With the definition of focal points, we also have a natural way to graphically solve optical problems. Any ray drawn horizontally from the left side of the interface will pass through the focal point on the right. Any ray going through the focal point on the left will go horizontally on the right. The following figure illustrates this.



The magnification of the image is the ratio of the heights h_o to h_i.



Since we are using the small angle approximation, we have Snell's law

$$n_1\theta_1 = n_2\theta_2$$

which can be rewritten

$$n_1\left(\frac{h_o}{s_o}\right) = n_2\left(\frac{h_i}{s_i}\right).$$

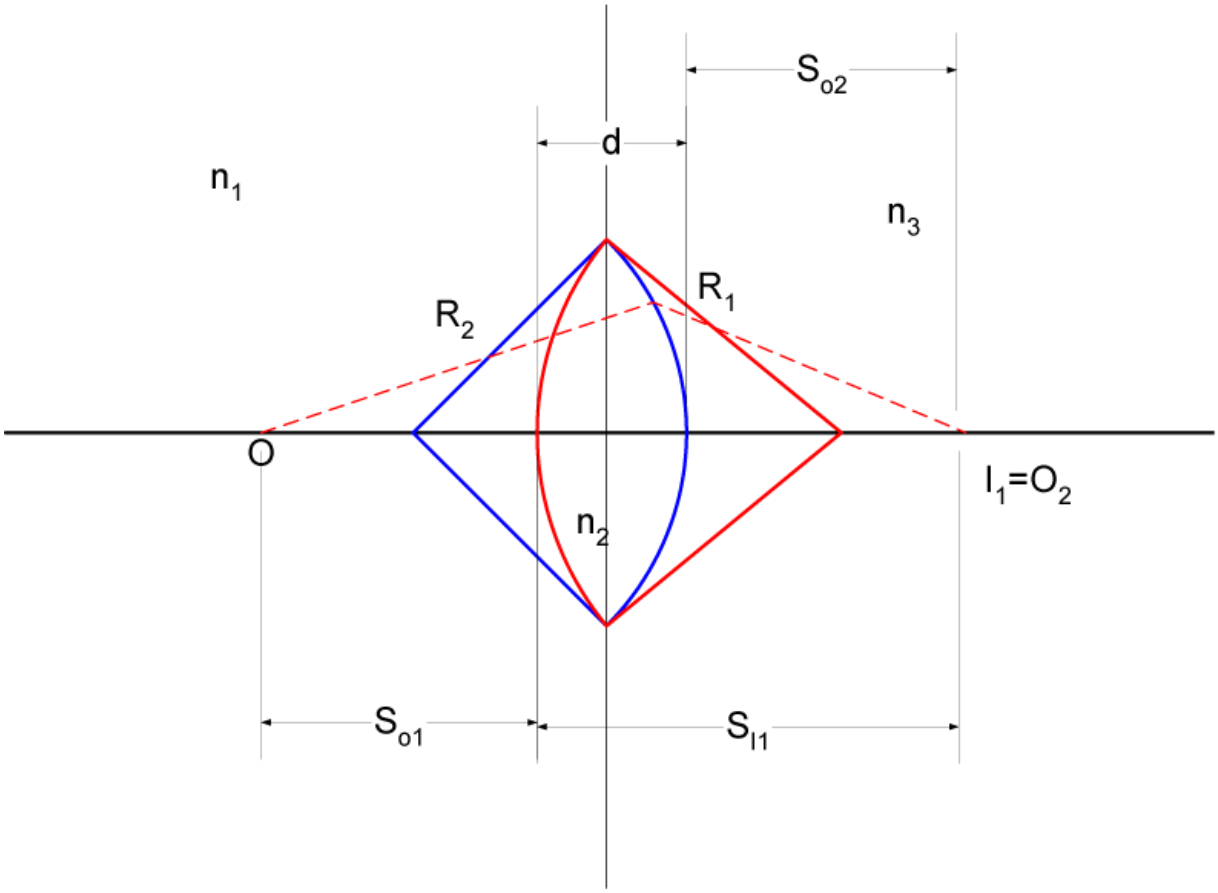
So we write that the magnification is

$$m = \frac{h_i}{h_o} = -\frac{n_1 s_i}{n_2 s_o}.$$

The negative sign is introduced to capture the fact that the image is inverted. It is worth pointing out that in our diagram above, the image is real, because the actual light rays pass through it.

Thin Lens

Now we can turn to the case of lenses. A lens can be considered the combination of two spherical interfaces. To solve an optical problem using multiple interfaces or lenses, one considers each one by one. For example one finds the image created by the first surface and then uses it as the object of the second surface.



Consider a lens of thickness d as shown in the drawing. At interface 1 (coloured red in the drawing) we have

$$\frac{n_1}{s_{o_1}} + \frac{n_2}{s_{i_1}} = \frac{n_2 - n_1}{R_1}$$

For surface 2 (coloured blue) the image of the the first surface becomes the object of the second. Note the sign of s_{o_2} is negative so that

$$s_{o_2} = d - s_{i_1}$$

Thus

$$\frac{n_2}{s_{o_2}} + \frac{n_3}{s_{i_2}} = \frac{n_3 - n_2}{R_2}$$

becomes

$$\frac{n_2}{d - s_{i_1}} + \frac{n_3}{s_{i_2}} = \frac{n_3 - n_2}{R_2}$$

Now add the equations

$$\frac{n_1}{s_{o_1}} + \frac{n_2}{s_{i_1}} + \frac{n_2}{d - s_{i_1}} + \frac{n_3}{s_{i_2}} = \frac{n_2 - n_1}{R_1} + \frac{n_3 - n_2}{R_2}$$

Now we take the thin lens case,that is $d \to 0$

$$\frac{n_1}{s_o} + \frac{n_3}{s_i} = \frac{n_2 - n_1}{R_1} + \frac{n_3 - n_2}{R_2}$$

That equation would work for making prescription swim goggles for example, however most of the time $n_1 = n_3$ (namely air for eyeglasses). So making that the case we get

$$\frac{n_1}{s_o} + \frac{n_1}{s_i} = (n_2 - n_1)\left[\frac{1}{R_1} - \frac{1}{R_2}\right]$$

which in the case of air (n=1) is

$$\frac{1}{s_o} + \frac{1}{s_i} = (n_l - 1)\left[\frac{1}{R_1} - \frac{1}{R_2}\right].$$

That is The lensmaker's formula(where n_l is the index of refraction of the lens material)Now we can find the foci

$$f_i = \lim_{s_i \to \infty}$$

$$f_o = \lim_{s_o \to \infty}$$

Which we see from the lensmaker's formula must be the same so lets drop the subscripts:

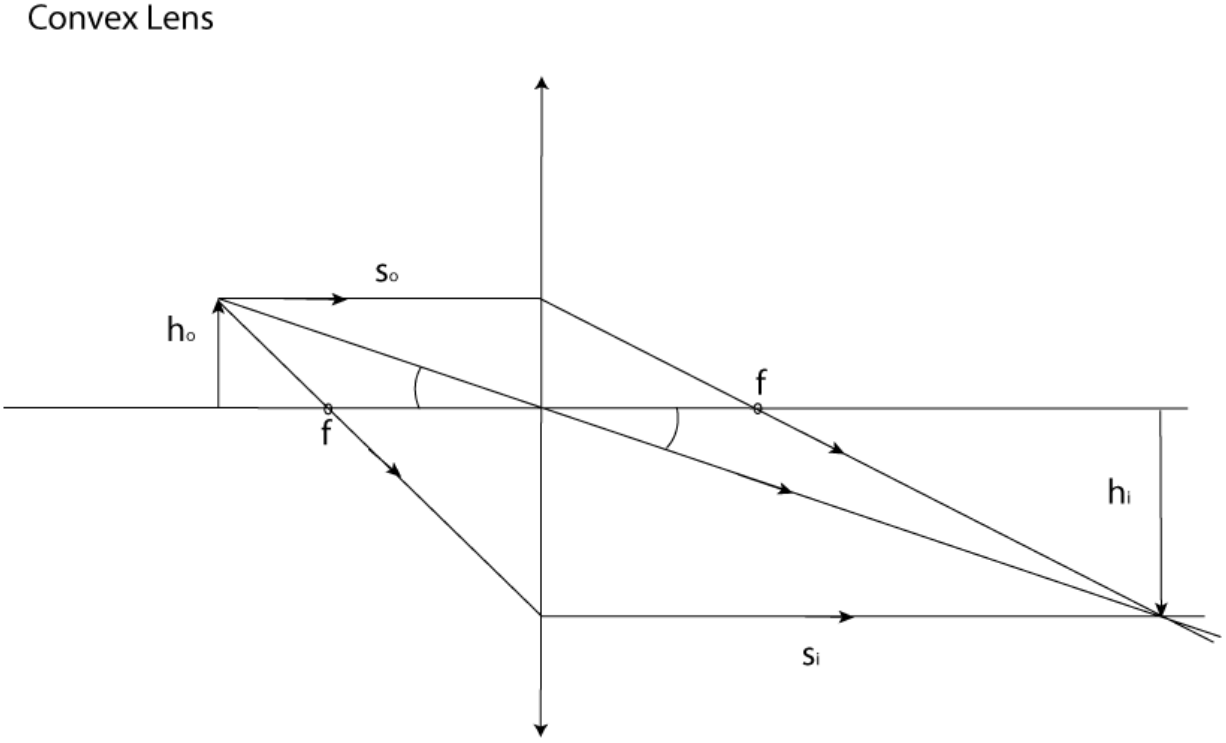
$$\frac{1}{f} = (n_l - 1)\left[\frac{1}{R_1} - \frac{1}{R_2}\right]$$

and

$$\frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{f}$$

which is the Gaussian Lens Formula

A convex lens will have a positive focal length f.



We can use the same equation as before for the magnification,

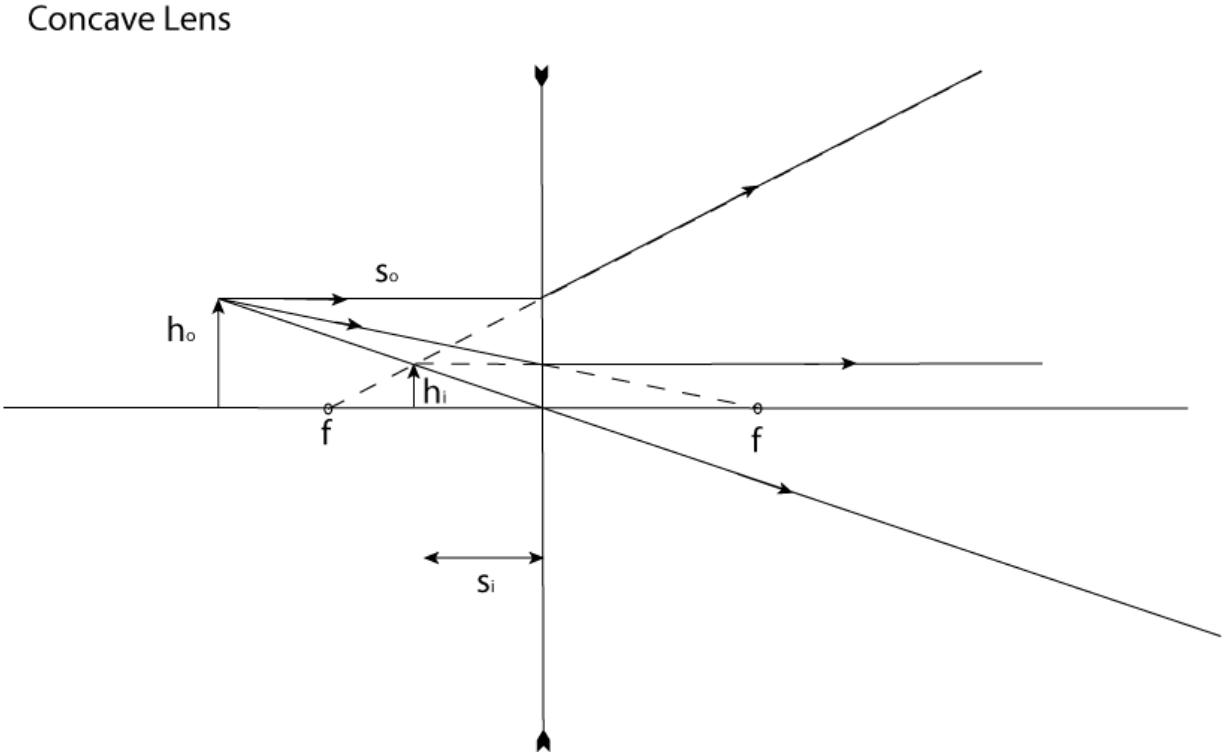
$$m = \frac{h_i}{h_o} = -\frac{n_1 s_i}{n_2 s_o}.$$

but now note that $n_1 = n_2$ so the equation becomes

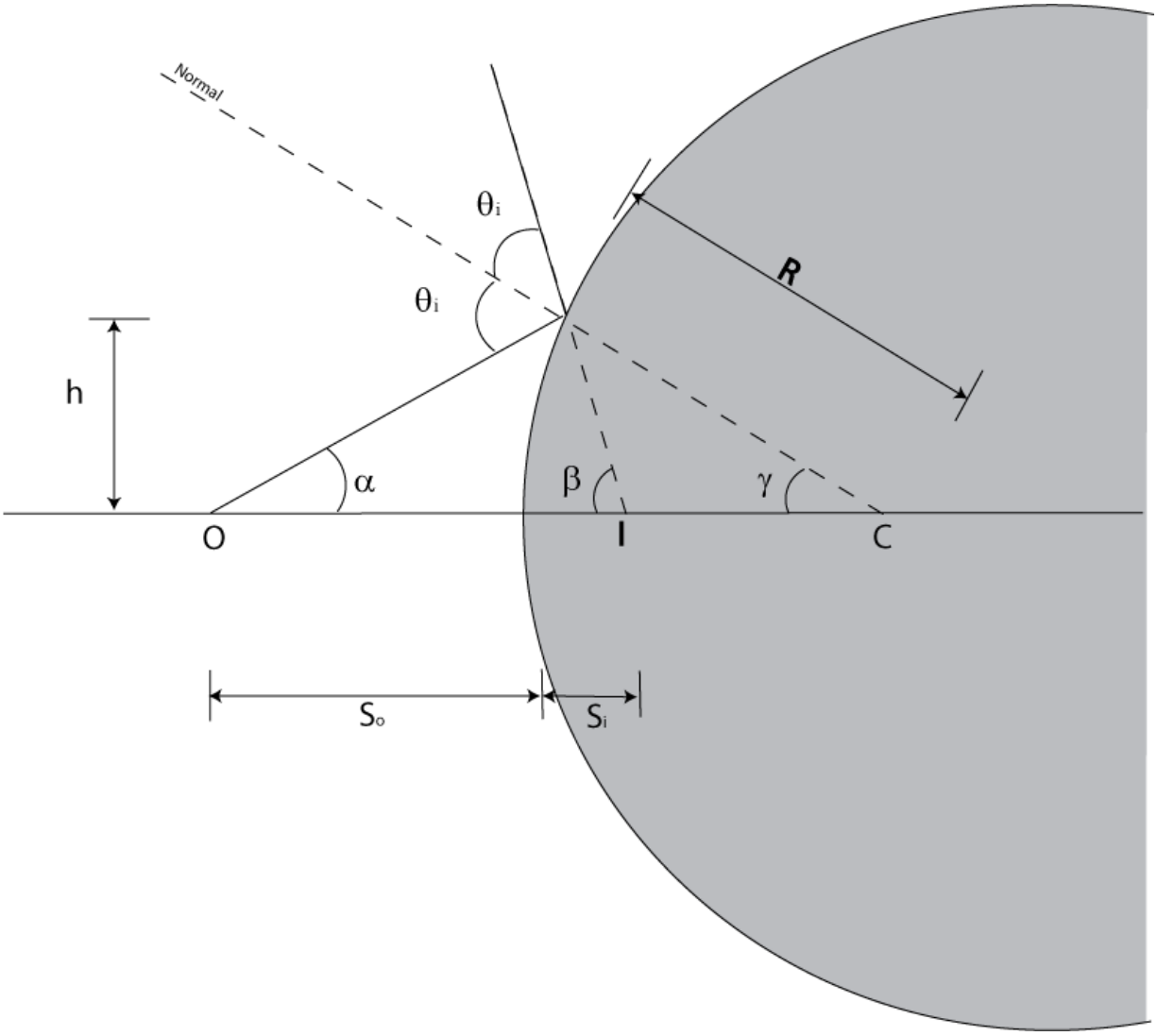
$$m = \frac{h_i}{h_o} = -\frac{s_i}{s_o}.$$

You can examine the figure above to verify that this is true.

We can also consider a concave lens which has a negative focal length. Notice that in this case the image is upright and virtual. Notice that in this case s_i is negative.



Mirrors



Notice that in Mirrors we get a virtual image to the right of the surface. Thus for mirrors we say that s_i is positive to the left of the mirror. This allows to retain correspondence between s_i being negative and an image being virtual.

Again we use the small angle approximation. By inspection of the figure we see that

$$2\theta_i = \alpha + \beta$$

and

$$\theta_i = \alpha + \gamma.$$

Now we multiply the second equation by two and subtract the first equation from it and we get:

$$2\alpha + 2\gamma - \alpha - \beta = 0$$

or

$$\alpha - \beta = -2\gamma.$$

Using the small angle approximation we see that this is

$$\frac{h}{s_o} + \frac{h}{s_i} = \frac{-2h}{r}$$

where I have used the fact that s_i is negative to the right of the mirror. So I can write the mirror equation as

$$\frac{1}{s_o} + \frac{1}{s_i} = \frac{-2}{R}$$

or

$$\frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{f}$$

where for a mirror $1/f = -2/R$

Eyeglasses

First a little nomenclature. Optometrists (and opthamologists) use the dioptric power measured in diopters. A diopter is $1/f$ where f is measured in meters. The focal length of lenses in contact is

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}$$

or using dioptric power

$$D = D_1 + D_2$$

A "normal" eye will focus an object at infinity onto the retina with the lens relaxed. As objects come closer the lens of the eye is made to bulge to keep the object in focus. A nearsighted person (such as certain Phys 201 prof's) has an eye that focuses the object at infinity in front of the retina. This person has a far point beyond which things are focused incorrectly. By using a lens that casts the object at infinity onto the far point this problem can be corrected. Say someone's far point is at 2m. In this case we want

$$\frac{1}{f} = \frac{1}{s_o} + \frac{1}{s_i}$$

$$\frac{1}{f} = \frac{1}{\infty} + \frac{1}{-2}$$

a corrective lens with $f = -2m$ or $D = -\frac{1}{2}$Diopters. Notice that the desired focal length is the far point (with the correct choice of sign)Of course in practice we need to take into account the distance between the glasses and the eye, but the above is valid for a contact lens. In the case of eyeglasses you need to subtract off the distance between the lense and the eye. For example if the distance between the eye and the glasses is 2cm then the above becomes

$$\frac{1}{f} = \frac{1}{\infty} + \frac{1}{-1.98}$$

A far sighted person focuses objects at infinity behind the retina. Their eye can accommodate the object at infinity but they have trouble with nearby points. A normal eye should be able to focus objects as close as 25cm. A far sighted person can focus to a near point that is greater than 25cm and so we need to take the 25cm point and make it look like it is at the near point. Say some one has a near point of 125cm then

$$\frac{1}{f} = \frac{1}{-1.25} + \frac{1}{.25}$$

which is $+3.2$Diopters. Of course again we should correct for the distance between the glasses and the eye, which if that is 2cm makes the equation:

$$\frac{1}{f} = \frac{1}{-1.23} + \frac{1}{.23}$$

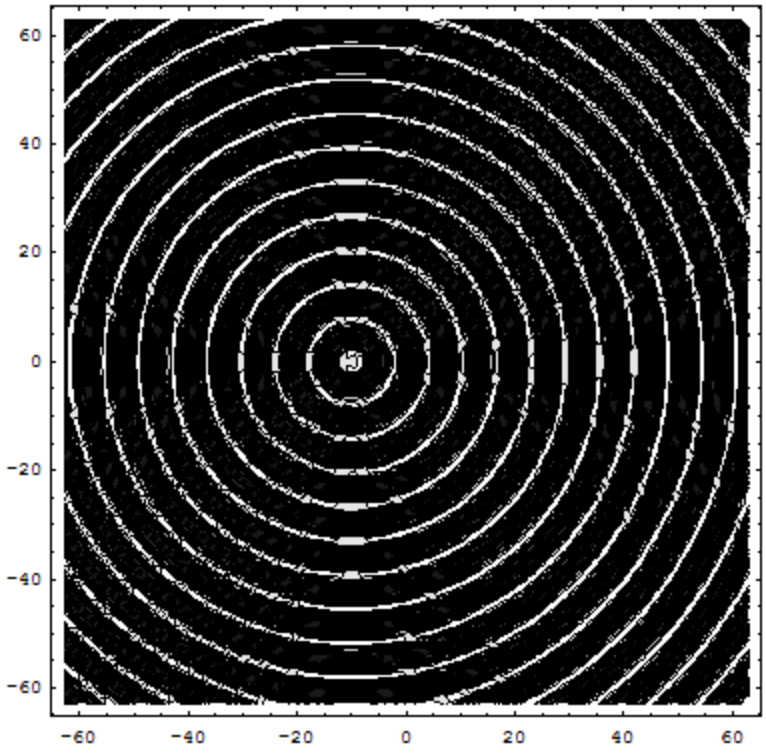
Finally their are old guys like me who need bifocals. As one ages, the eye lens becomes more rigid and it is harder and harder for it to deform and focus on close in objects. Then if you are also nearsighted, then you have to resort to bifocals.

Two Source Interference
We examine interference from two coherent sources.

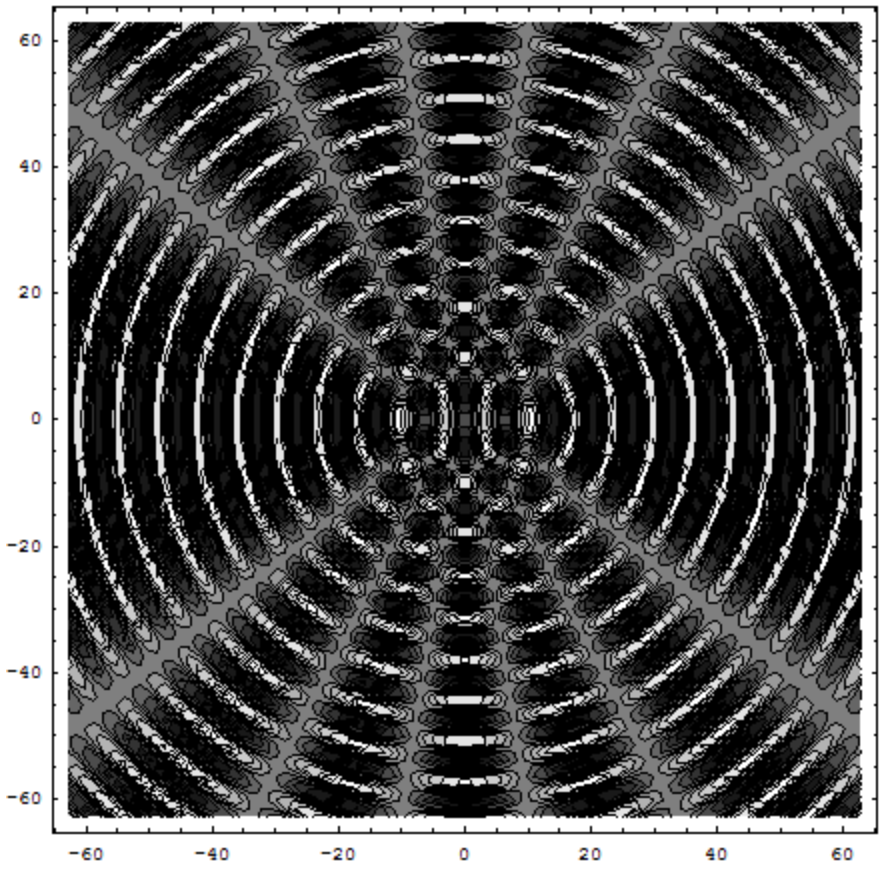
Interference

Waves on a pond:

Think of when you drop a pebble into a pond, you will see circular waves eminate from the point where you dropped the pebble.



When you drop two pebbles side by side you will see a much more complicated pattern:



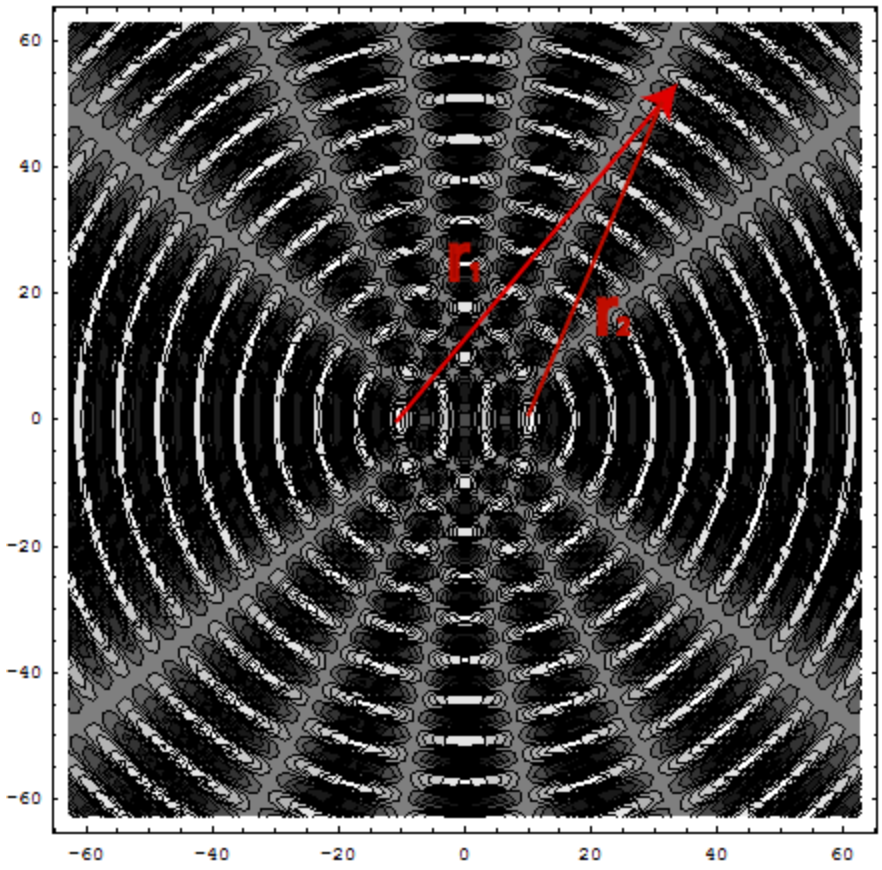
Likewise with electromagnetic waves, you can get interesting interference phenomena when waves eminate from two point sources.

Two Point Sources

Lets take a particular example of two point sources separated by a distance d. The waves emitted by point source are spherical and thus can be written

$$\vec{E} = \frac{\vec{E}_0}{r} \cos\left(kr - \omega t\right)$$

To make the problem easier we will make the k's the same for the two sources. Also lets set the E_0's to be the same as well.



The the only difference in the waves will be the r's, that is

$$\vec{E}_1 = \frac{\vec{E}_0}{r_1} \cos\left(kr_1 - \omega t\right)$$

$$\vec{E}_2 = \frac{\vec{E}_0}{r_2} \cos\left(kr_2 - \omega t\right)$$

Now there is a slightly subtle point here that is important to understand. In the denominator it is sufficient to say that $r_1 \approx r_2$ and just call it r. We assume that we are far enough away that the differences between r_1 and r_2 are too small to matter. However this is not true in the argument of the harmonic function. There, very small differences between r_1 and r_2 can have a big effect. So lets define $r_1 = r_2 = R$

$$
\begin{aligned}
I &\propto \left\langle \left(\tfrac{E_0}{R}\cos\left(kr_1-\omega t\right) + \tfrac{E_0}{R}\cos\left(kr_2-\omega t\right)\right)^2 \right\rangle_T \\
&= \left\langle \tfrac{E_0^2}{R^2}\cos^2\left(kr_1-\omega t\right)\right\rangle_T + \left\langle \tfrac{E_0^2}{R^2}\cos^2\left(kr_2-\omega t\right)\right\rangle_T \\
&\quad + \left\langle 2\tfrac{E_0^2}{R^2}\cos\left(kr_1-\omega t\right)\cos\left(kr_2-\omega t\right)\right\rangle_T \\
&= \tfrac{1}{2}\tfrac{E_0^2}{R^2} + \tfrac{1}{2}\tfrac{E_0^2}{R^2} + \\
&\quad + \left\langle 2\tfrac{E_0^2}{R^2}\cos\left(kr_1-\omega t\right)\cos\left(kr_2-\omega t\right)\right\rangle_T
\end{aligned}
$$

Now to evaluate the final term we use

$$\cos\left(\theta-\varphi\right) = \cos\theta\cos\varphi + \sin\theta\sin\varphi$$

and write So we have

$$
\begin{aligned}
I &\propto \tfrac{1}{2}\tfrac{E_0^2}{R^2} + \tfrac{1}{2}\tfrac{E_0^2}{R^2} + 2\tfrac{E_0^2}{R^2}\tfrac{1}{2}\cos k\Delta r \\
&= \tfrac{1}{R^2}(E_0^2 + E_0^2\cos k\Delta r) \\
&= \tfrac{1}{R^2}E_0^2(1+\cos k\Delta r)
\end{aligned}
$$

Clearly I will be a maximum when the cosine is = +1

$$k\Delta r = 2n\pi \qquad n = 0, 1, 2\ldots$$

$$\frac{2\pi}{\lambda}\Delta r = 2n\pi$$

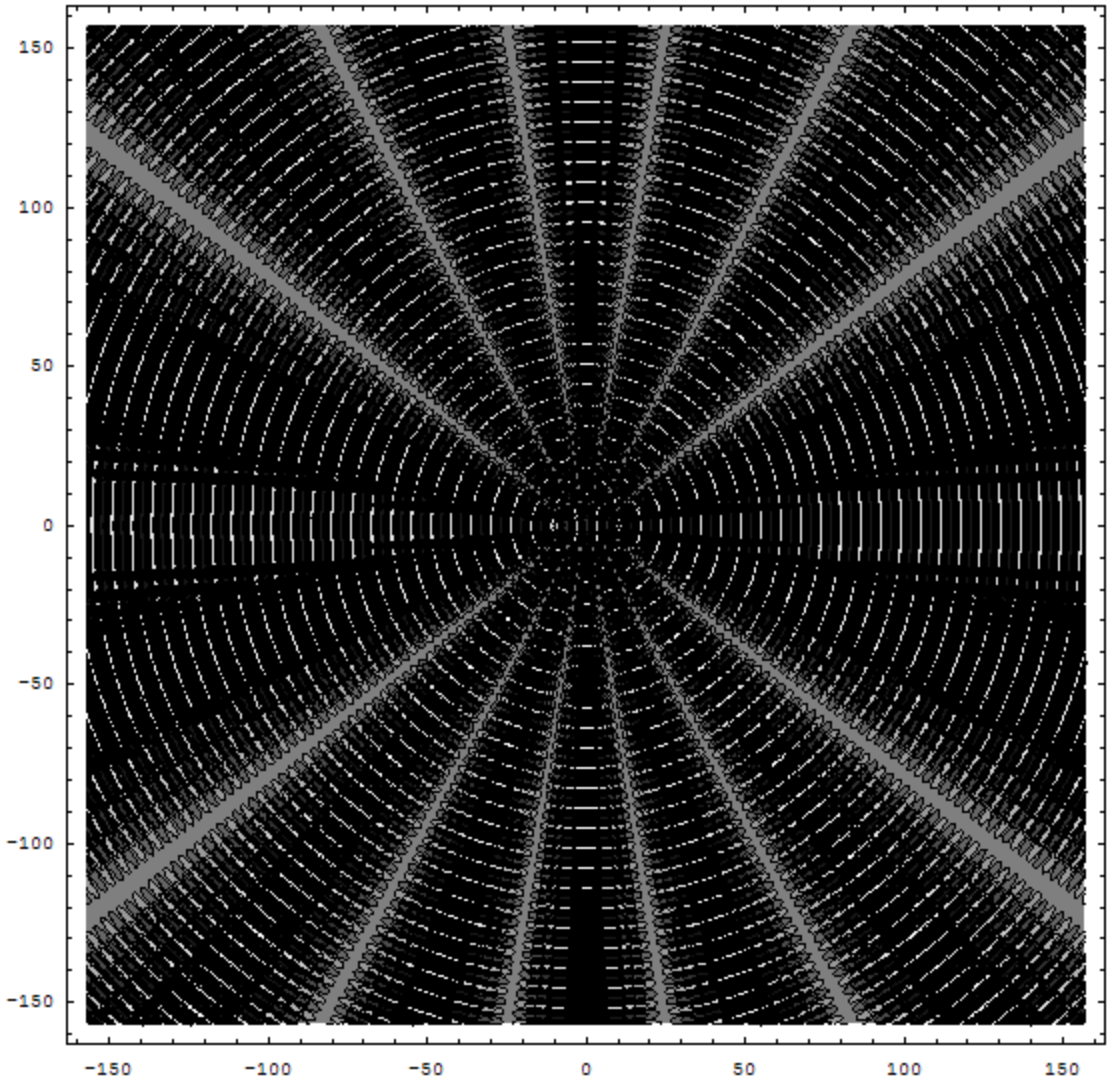
$$\Delta r = n\lambda$$

There will be a minimum when the cosine is = -1

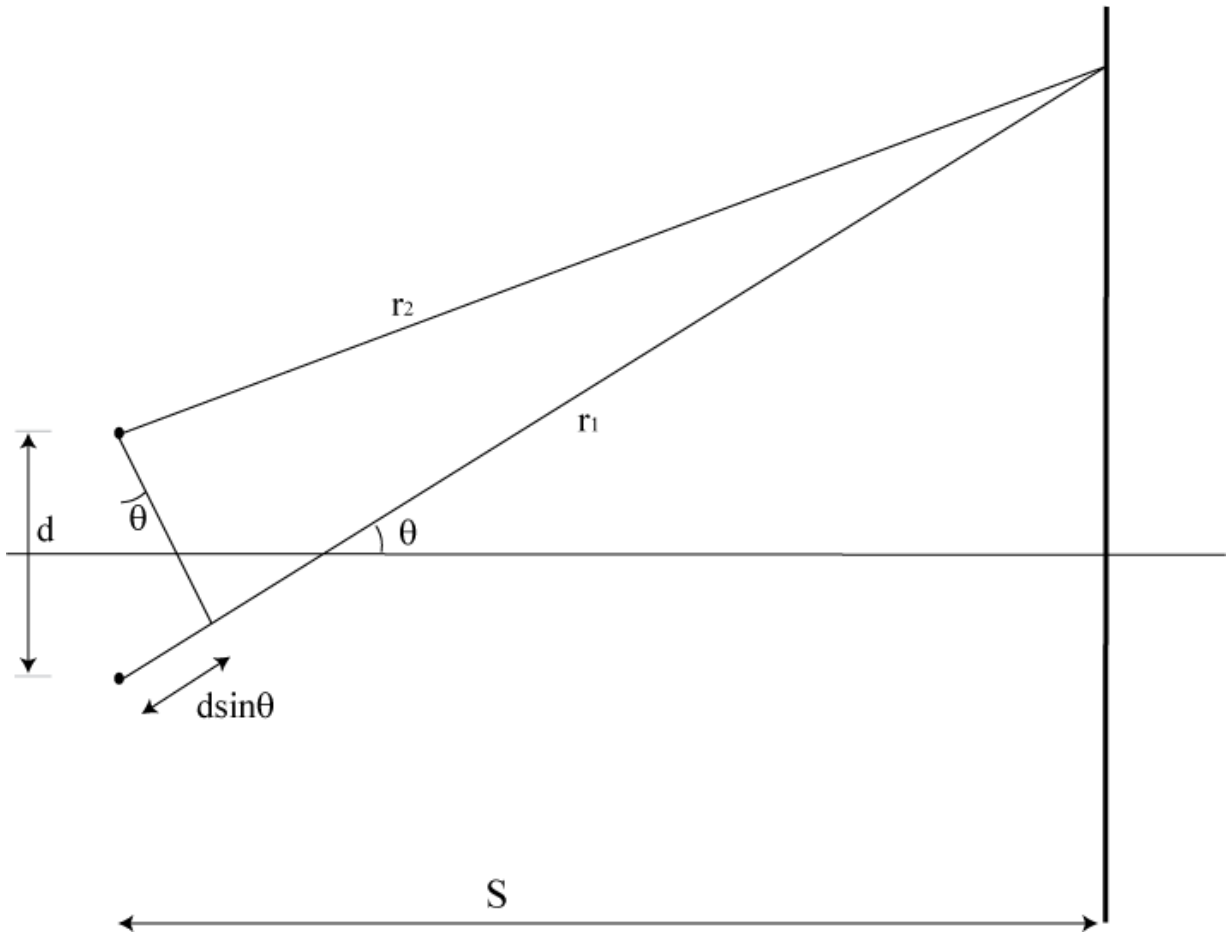
$$k\Delta r = n\pi \qquad n = 1, 3, 5\ldots$$

$$\Delta r = \frac{n\lambda}{2} \qquad n = 1, 3, 5\ldots$$

So you get light and dark bands which are called interference fringes.To reiterate; we have two rays of light eminating from two point sources. We have looked at the combined wave at some point, a distance r_1 from the first source and a distance from the second source. In that case we find that the intensity is proportional to $\frac{1}{R^2}E_0^2(1+\cos k\Delta r)$. To make things easier we can redefine E_0 to be the amplitude of the waves at the point under consideration, that is $I=\varepsilon_0 cE_0^2(1+\cos k\Delta r)$. Or we can say $I_0=\varepsilon cE_0^2/2$ and write $I=2I_0(1+\cos k\Delta r)$.



Say we place a screen a distance S away from the two sources:



In this case we see that

$$\Delta r = d \sin\theta$$

So we have maxima at

$$\Delta r = n\lambda = d \sin\theta.$$

The angle between two maxima is given by

$$\sin\theta_{n+1} - \sin\theta_n = \frac{\lambda}{d}$$

or for small θ

$$\Delta\theta = \frac{\lambda}{d}$$

Notice how when the sources are moved far apart the effect maxima become very close together so the screen appears to be uniformly illuminated. If a screen is placed a distance S away the maxima on the screen will occur such that

$$d \sin \theta = n\lambda$$

but in the small angle limit

$$\sin \theta = \tan \theta = \frac{y}{S}$$

which implies

$$y = \frac{n\lambda S}{d}$$

likewise minima will occur at

$$y = \frac{n\lambda S}{2d} \qquad n = 1, 3, 5\ldots$$

using $\cos \theta = 2 \cos^2 \frac{\theta}{2} - 1$ we can rewrite

$$I = 2I_0(1 + \cos k\Delta r)$$

as

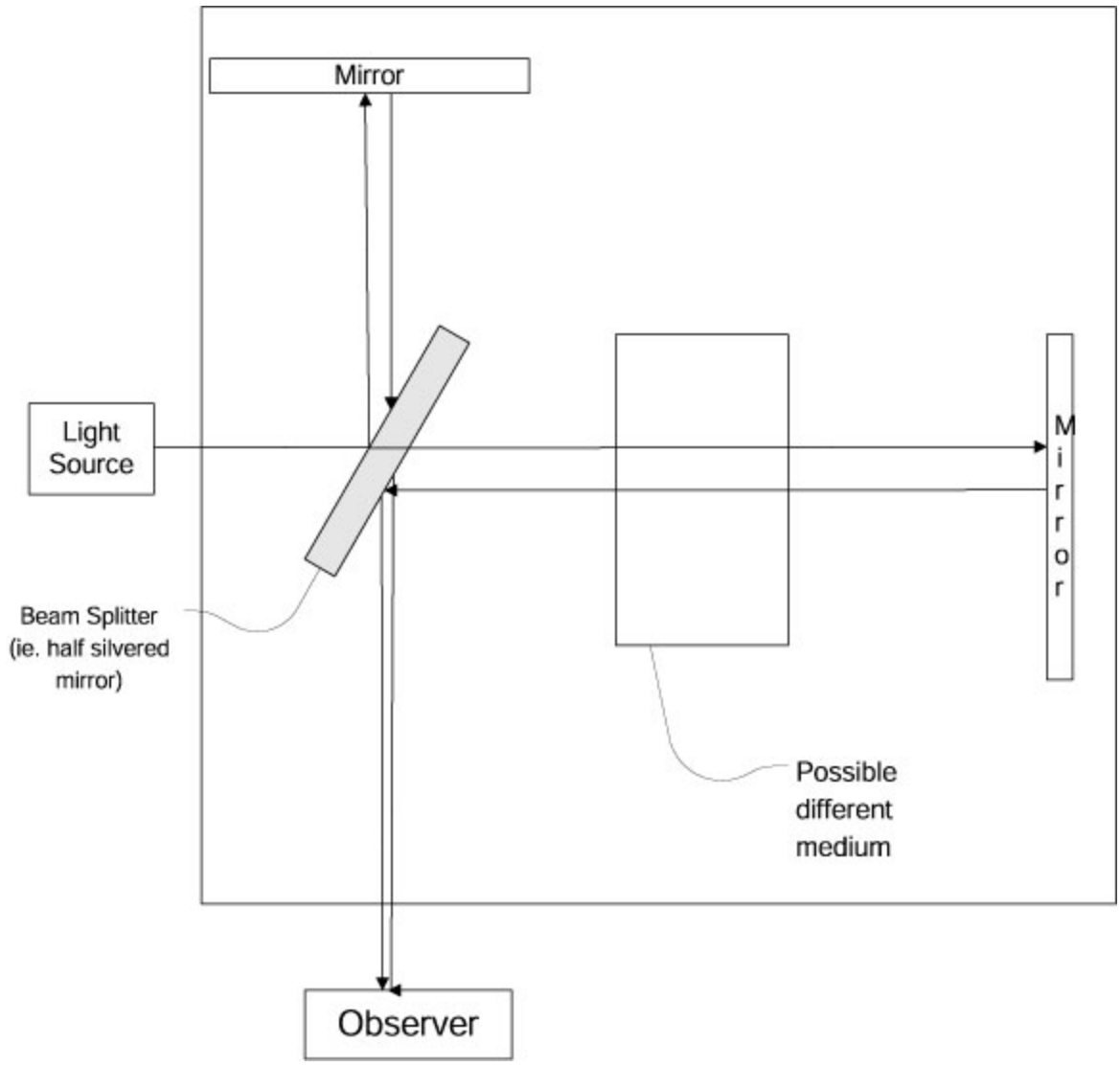
$$I = 4I_0 \cos^2 \frac{k\Delta r}{2}$$

Young's Double Slit

Young's double slit.is an excellent example of two source interference. The equations for this are what we worked out for two sources above. Interference is an excellent way to measure fine position changes. Small changes in Δr make big observable changes in the interference fringes.

Michelson Interferometer

A particularly useful example of using interference is the Michelson interferometer. This can be used to measure the speed of light in a medium, measure the fine position of something, and was used to show that the speed of light is a constant in all directions.



When Δr, the path length difference in the two arms is $\Delta r = n\lambda$ then the rays of light in the traveling down the center of the apparatus will interfere constructively. As you move off axis the light travels slightly different lengths and so you get rings of interference patterns. If you have set up the

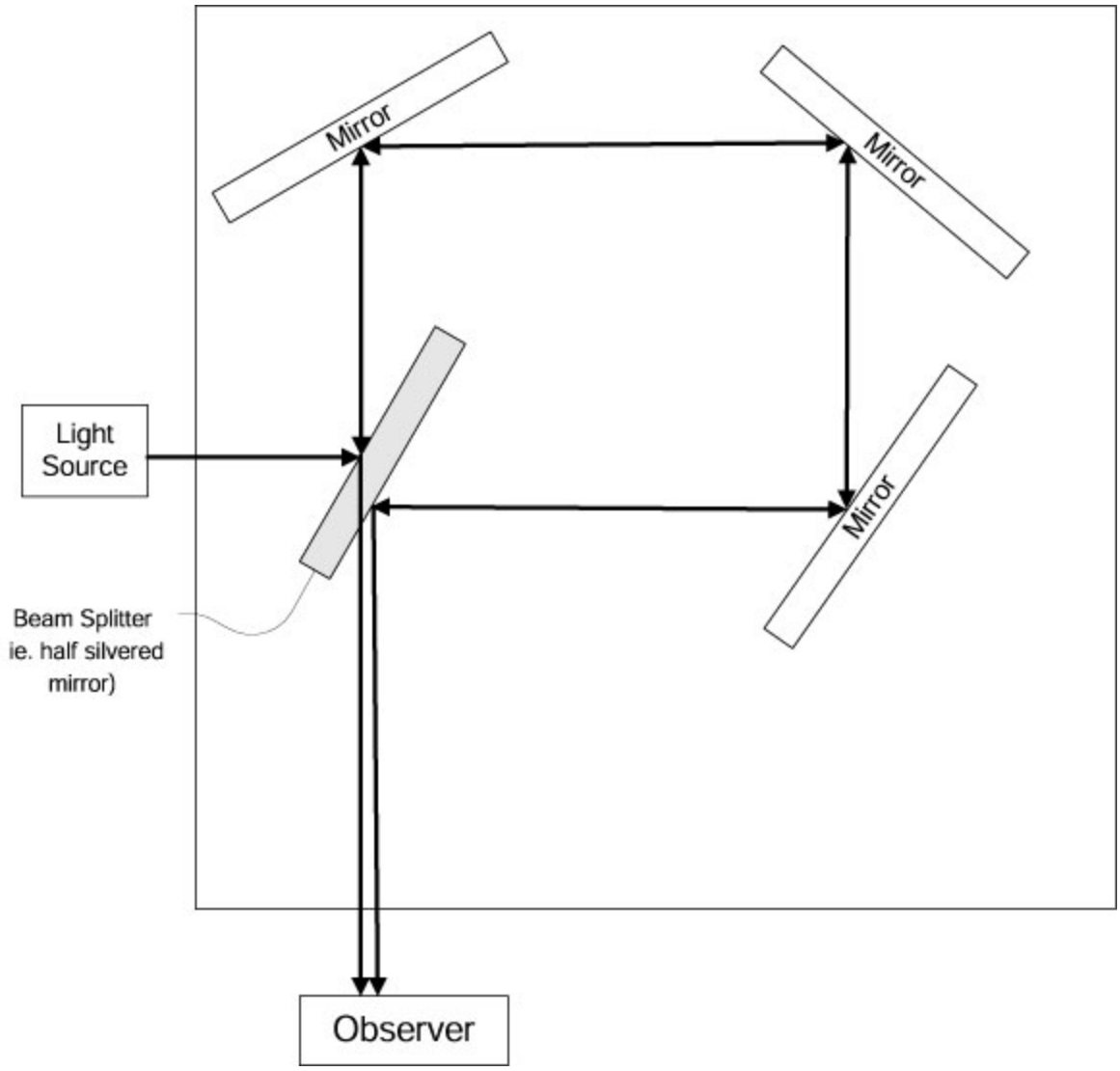
apparatus so that $\Delta r = n\lambda$ and then move one of the mirrors a quarter wavelength then $\Delta r = n\lambda + \frac{1}{2}\lambda$ and you get destructive interference of the central rays. Thus you can easily position things to a fraction of a micron with such a set up.

What really matters is the change in the optical pathlength. For example you could introduce a medium in one of the paths that has a different index of refraction, or different velocity of light. This will change the optical pathlength and change the interference at the observer. Thus you can measure the velocity of the light in the introduced medium.

Michelson and Morely used this technique to try to determine if the speed of light is different in different directions. They put the whole apparatus on a rotating table and then looked for changes in the interference fringes as it rotated. They saw no changes. In fact they went so far as to wait to see what happened as the earth rotated and orbited and saw no changes. They thus concluded that the speed of light was the same in all directions (which nobody at the time believed, even though that is the conclusion you draw from Maxwell's equations.)

Ring Gyroscope

Another application of interference is a a gyroscope, ie. as device to measure rotations.



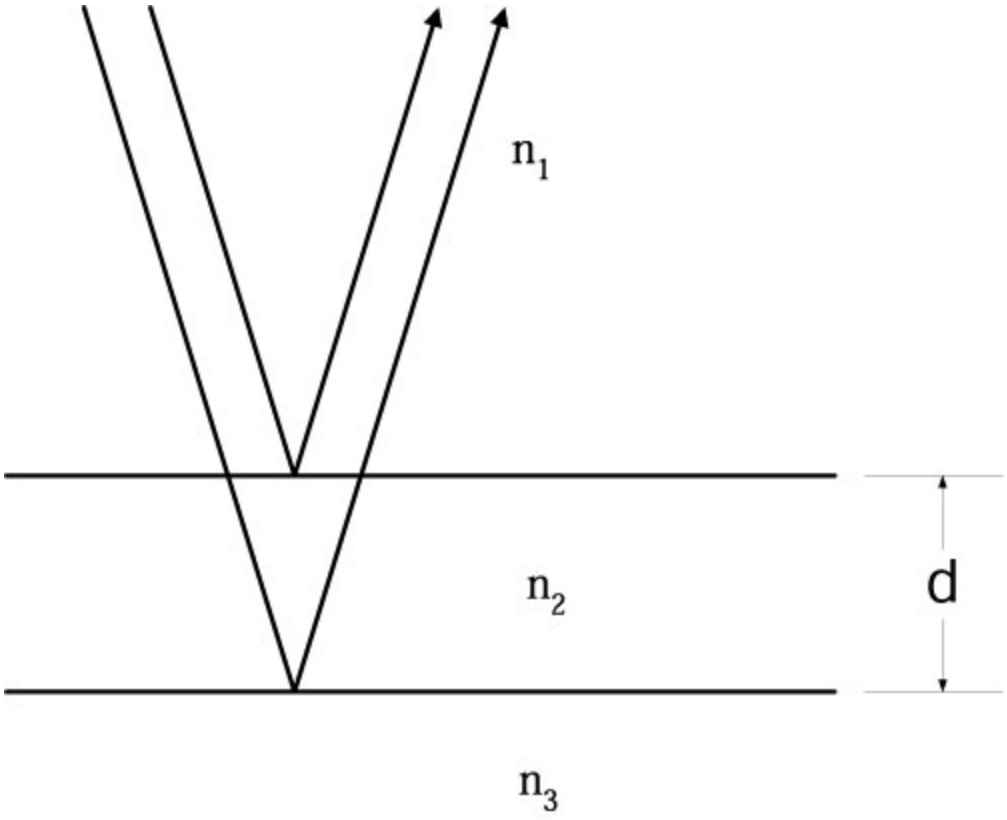
If the apparatus is rotating, then the pathlengths are different in different directions and so you can use the changes in the interference patterns to measure rotations. This is in fact how gyroscopes are implemented in modern aircraft.

Thin Film Interference

Interference

Thin Films

Suppose there is a very thin film of dielectric and light is incident on it normally. Lets consider single reflections. (We make the small angle of incidence approximation)



We will assume $n_3 > n_2 > n_1$. The physical path length difference of the reflected light is $\Delta r = 2d$. We will get maxima in the interference when:

$$\Delta r = 2d = m\lambda_2 \qquad m = 1, 2, 3\ldots$$

where λ_2 is the wavelength in the film. Now

$$\lambda_i \nu_i = c/n_i.$$

In our example we have $\nu_1 = \nu_2 = \nu_3$, that is the frequency does not change moving between the media. So we have

$$\lambda_1 n_1 = \lambda_2 n_2 = \lambda_3 n_3.$$

Thus constructive interference will happen when

$$\begin{gathered}
\lambda_2 = \lambda_1 \tfrac{n_1}{n_2} \\
2d = m\lambda_2 \quad m = 1, 2, 3.. \\
2d = m\lambda_1 \tfrac{n_1}{n_2} \\
2d = m\lambda_{air} \tfrac{n_{air}}{n_{film}} \\
2d = m\lambda_1 \tfrac{1}{n_{film}} \\
(2d)n_{film} = m\lambda_{air} \quad m = 1, 2, 3...
\end{gathered}$$

where $n_{film} = n_2$. Destructive interference will happen when

$$(2d)n_{film} = m\lambda_{air}/2 \quad m = 1, 3, 5...$$

When destructive interference occurs then that value of λ is not reflected. Note that this is a function of both d and λ. The next effect is that different colours of light get reflected at different thicknesses of the film. This is why soap films or oil films on water give rainbow effects.

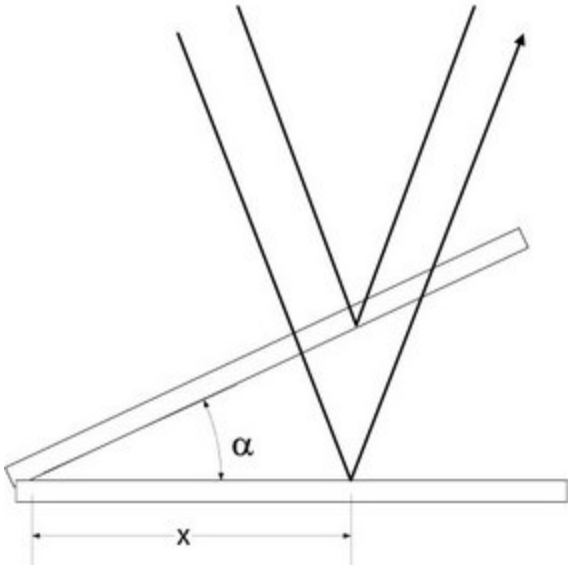
Note I have assumed that $n_3 > n_{film} > n_{air}$ in the above, where n_3 is the material that the film sits upon.

Consider an interface between two materials with indices of refraction n_1 and n_2. If $n_2 > n_1$. Then lets examine what happens to the phase of an electromagnetic wave upon reflection. For a transverse electric field, there is a phase change of π. For the transverse magnetic field (or $E_{\|}$) there is not, if the light ray is close to the normal. However if $n_1 > n_2$ then and the situation is reversed and the transverse electric field does not undergo a phase change and the transverse magnetic field does. In the example above, their will be no relative phase change between the rays in either case. Either

both will change by π or neither will change, depending on the orientation of the E field.

DoublePlateInterference.xhtml

Double Plate Interference



Assume that the normally incident light bounces off the bottom of the top glass and the top of the bottom glass. In this case there is a phase change between the two cases and the conditions of constructive and destructive interference flip.

So constructive interference occurs when

$$2d = n\lambda/2 \qquad n = 1, 3, 5\ldots$$

$$2x\alpha = n\lambda/2$$

or the position of the maxima are

$$x_{max} = \frac{n\lambda}{4\alpha} \quad n = 1, 3, 5\ldots$$

Destructive interference will occur at

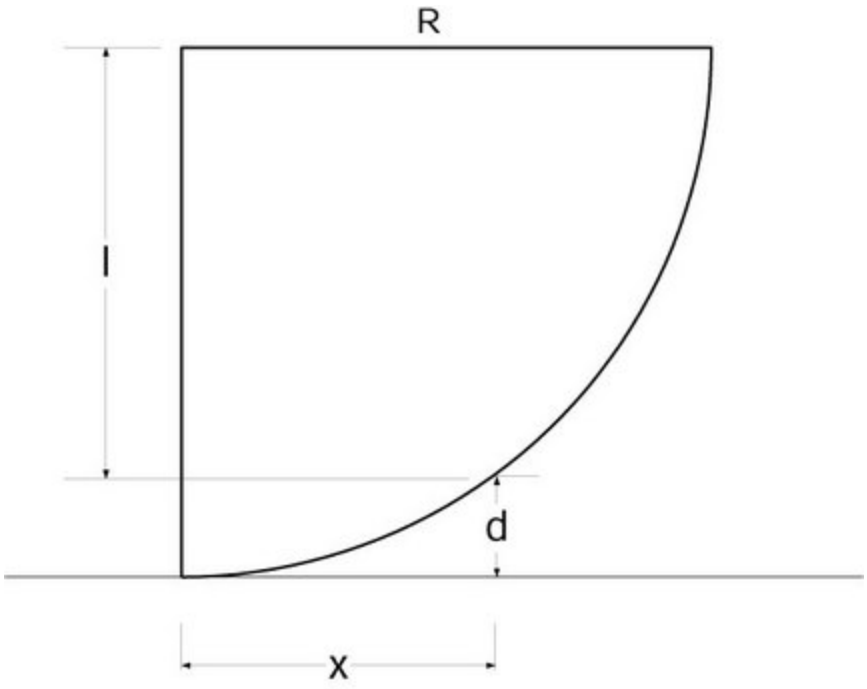
$$2d = n\lambda \quad n = 1, 2, 3, \ldots$$

$$x_{min} = \frac{n\lambda}{2\alpha} \quad n = 1, 2, 3\ldots$$

NewtonsRings.xhtml

Newton's Rings

Consider a flat surface in contact with a spherical surface.



$$x^2 + l^2 = R^2$$

$d = R - l$

$$\begin{aligned} d^2 &= R^2 + l^2 - 2Rl \\ &= R^2 + R^2 - x^2 - 2Rl \\ &= 2R^2 - x^2 - 2R(R - d) \\ &= 2R^2 - x^2 - 2R^2 + 2Rd \end{aligned}$$

or

$$x^2 = 2Rd - d^2$$

For small d we have $d^2 \to 0$ so

$$x^2 \approx 2Rd$$

In this case there is no phase change at the spherical surface but there is at the flat surface. So again there is an offset in the phase and the condition of constructive interference is

$$2d = n\lambda/2 \qquad n = 1, 3, 5, \ldots$$

or

$$2d = x^2/R = n\lambda/2 \qquad n = 1, 3, 5, \ldots$$

For a spherical interface one observes rings of light and dark. This in fact is used to check the sphericity of the lenses.

Multi Source Inteference

N-Source Interference

Lets look back at two source interference again, this time using exponential notation: We will consider both sources to be in phase with the same amplitude. Also, just like before we define E_0 to be at the point where the interference is occurring, and approximate that it is the same for both waves. We have

$$\vec{E}_1 = \vec{E}_0 e^{i(kr_1 - \omega t)}$$

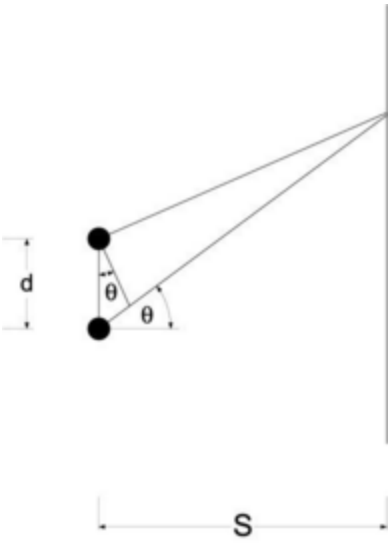
$$\vec{E}_2 = \vec{E}_0 e^{i(kr_2 - \omega t)}$$

or

$$\vec{E}_2 = \vec{E}_0 e^{i(k(r_1 + \Delta r) - \omega t)}$$

where

$$\Delta r = d \sin \theta$$



So I could write

$$\vec{E}_2 = \vec{E}_0 e^{i(kr_1 - \omega t)} e^{ik\Delta r}$$

or if $\delta = k\Delta r = kd\sin\theta$

$$\vec{E}_2 = \vec{E}_0 e^{i(kr_1 - \omega t)} e^{i\delta}$$

$$\vec{E}_2 = \vec{E}_1 (e^{i\delta})$$

So when we add the E fields together we get

$$\vec{E} = \vec{E}_1 + \vec{E}_2$$
$$= \vec{E}_0 e^{i(kr_1 - \omega t)} (1 + e^{i\delta})$$

Now if I consider the case of not 2 but N sources, all separated by a distance d, then this simply extends

$$\vec{E}_1 = \vec{E}_0 e^{i(kr_1 - \omega t)}$$

$$\vec{E}_2 = \vec{E}_0 e^{i(kr_2 - \omega t)}$$

$$\vec{E}_3 = \vec{E}_0 e^{i(kr_3 - \omega t)}$$

$$\vdots$$

$$\vec{E}_N = \vec{E}_0 e^{i(kr_N - \omega t)}$$

then $r_2 = r_1 + d\sin\theta$

$r_3 = r_1 + 2d\sin\theta$

and so on,

or $r_2 = r_1 + \Delta r$

$r_3 = r_1 + 2\Delta r$

So now when we add the E fields up we get

$$\vec{E} = \vec{E}_0 e^{i(kr_1 - \omega t)}(1 + e^{i\delta} + e^{2i\delta} + e^{3i\delta} + \ldots + e^{(N-1)i\delta})$$

or rewriting

Now following is a general property of geometric series:

$$\sum_{n=0}^{N-1} x^n = \frac{1 - x^N}{1 - x}$$

So now we get

$$\vec{E} = \vec{E}_0 e^{-i\omega t} e^{ikr_1} \frac{1 - e^{i\delta N}}{1 - e^{i\delta}}$$

or

$$\vec{E} = \vec{E}_0 e^{-i\omega t} e^{ikr_1} \frac{e^{i\delta N} - 1}{e^{i\delta} - 1}$$

or

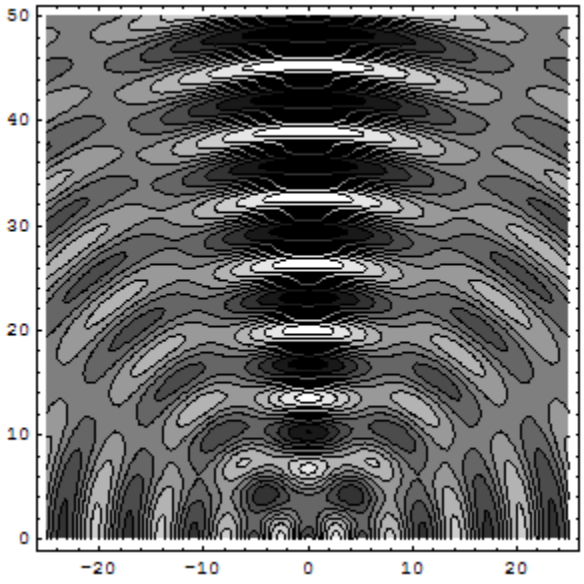
$$\vec{E} = \vec{E}_0 e^{-i\omega t} e^{ikr_1} \frac{e^{i\delta N/2}(e^{i\delta N/2} - e^{-i\delta N/2})}{e^{i\delta/2}(e^{i\delta/2} - e^{-i\delta/2})}$$

or

$$\vec{E} = \vec{E}_0 e^{-i\omega t} e^{i(kr_1 + (N-1)\delta/2)} \frac{\sin(\delta N/2)}{\sin(\delta/2)}$$

Now we can define $kR \equiv kr_1 + (N-1)\delta/2)$, which makes sense, this rephrases the equation in terms of the distance from the middle of the array of sources. So the equation becomes

$$\vec{E} = \vec{E}_0 e^{i(kR-\omega t)} \frac{\sin(\delta N/2)}{\sin(\delta/2)}$$



Interference pattern from three point sources

To find the irradiance, lets simplify things by taking the real part of this

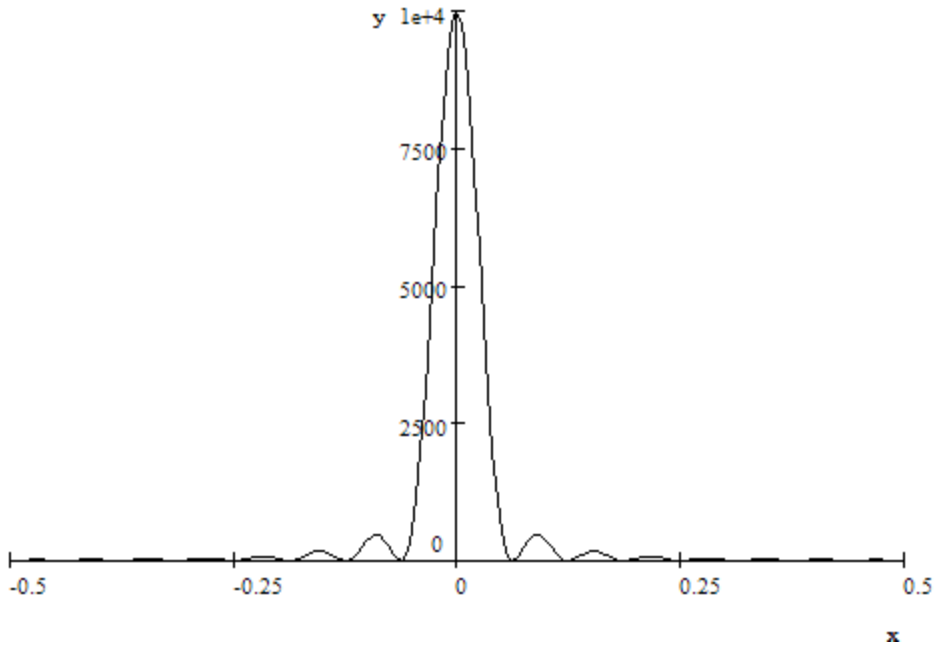
$$\vec{E} = \vec{E}_0 \cos(kR-\omega t) \frac{\sin(\delta N/2)}{\sin(\delta/2)}$$

Then

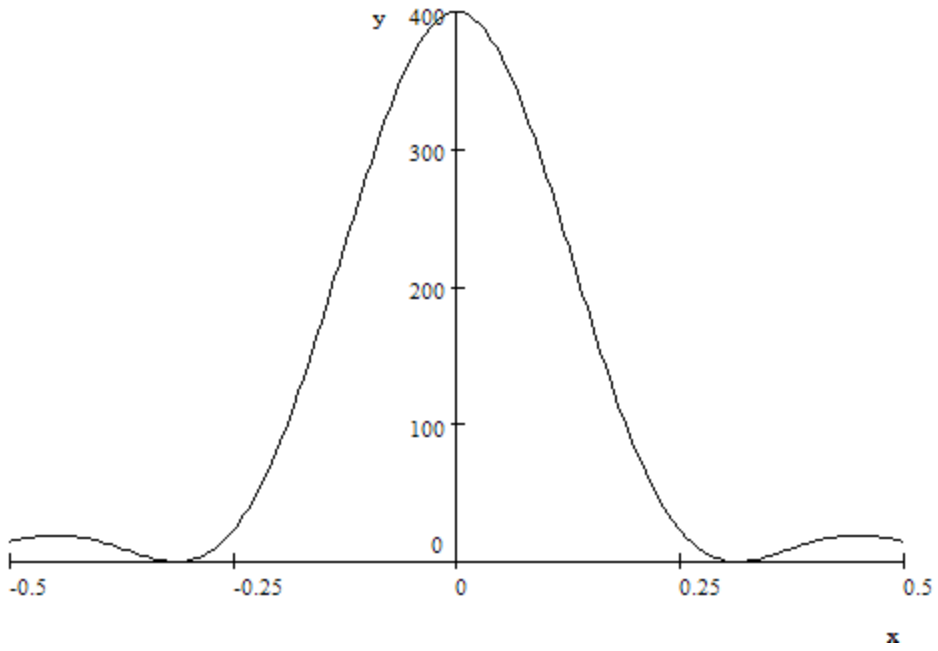
$$I = \varepsilon c < E^2 >= \varepsilon c \frac{E_0^2}{2} \frac{\sin^2(\delta N/2)}{\sin^2(\delta/2)}$$

or

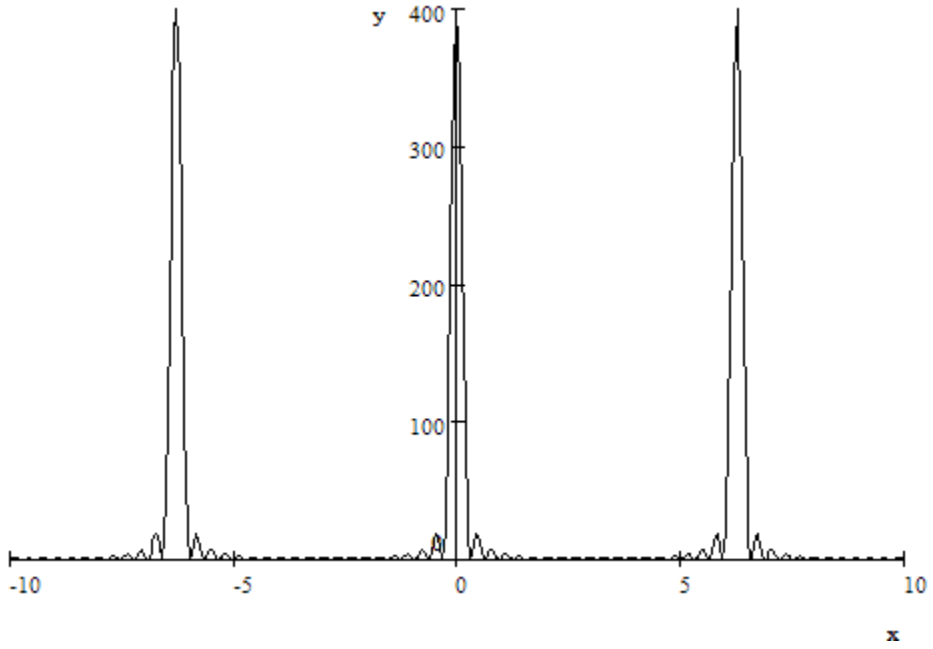
$$I = I_0 \frac{\sin^2(\delta N/2)}{\sin^2(\delta/2)}$$



Plot of $\frac{\sin^2(\delta 50)}{\sin^2(\delta/2)}$



Plot of $\frac{\sin^2(\delta 10)}{\sin^2(\delta/2)}$



Plot of $\frac{\sin^2(\delta 10)}{\sin^2(\delta/2)}$

Look at the plots, which show what the function $\frac{\sin^2(\delta N/2)}{\sin^2(\delta/2)}$ looks like for $N = 100$ and $N = 20$. The height of the first principle maximum is equal to N^2. This is because as

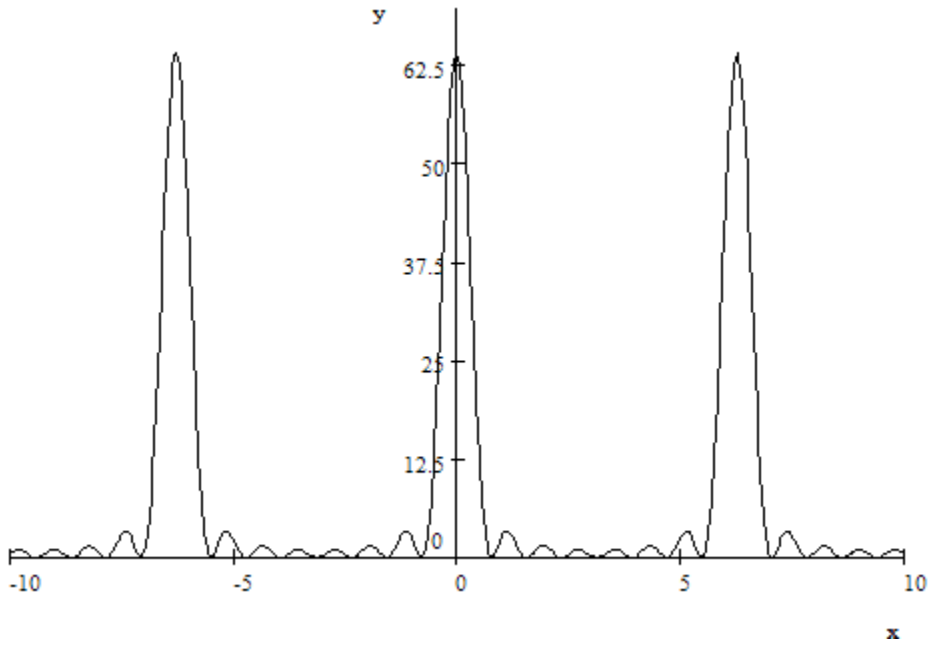
$\theta \to 0$ then $\delta \to 0$ (recall $\delta = kd\sin\theta$) Then

$$\lim_{\delta \to 0} \frac{\sin^2(\delta N/2)}{\sin^2(\delta/2)} \to \frac{(\delta N/2)^2}{(\delta/2)^2} = N^2$$

or at $\theta = 0$

$$I = N^2 I_0$$

It is also interesting to note that the first maxima become narrower as N becomes larger.



Plot of $\frac{\sin^2(\delta 4)}{\sin^2(\delta/2)}$

Principle maxima occur when

$$\delta/2 = n\pi$$

or

$$kd\sin\theta_{max} = 2n\pi \qquad n = 0, 1, 2, 3$$

or

$$\frac{2\pi}{\lambda} d\sin\theta_{max} = 2n\pi$$

or

$$\sin\theta_{max} = \frac{n\lambda}{d}$$

Minima occur when the numerator vanishes but the denominator does not:

$$N\delta/2 = n\pi \qquad n = 1, 2, 3\ldots \quad \frac{n}{N} \neq integer$$

$$kd\sin\theta = 2n\pi/N$$

$$\frac{2\pi}{\lambda}d\sin\theta = 2n\pi/N$$

or minima occur at

$$\sin\theta = \frac{n\lambda}{Nd} \qquad n = 1, 2, 3\ldots \quad \frac{n}{N} \neq integer$$

There are secondary maxima between the minima that are away from a principle maximum.

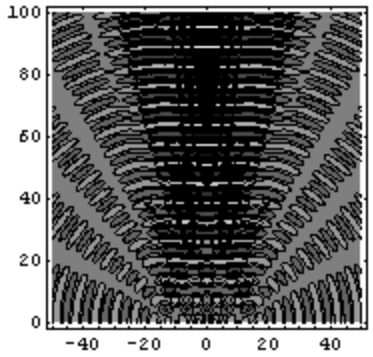
This gives us an insight into phase array radar and interferometric radio telescopes. Suppose you have a series or radar antennas in a row. Then you introduce a phase shift ε between each oscillator, then you get

$$\delta = kd\sin\theta + \varepsilon$$

and principle maxima will occur at

$$d\sin\theta_{max} = n\lambda - \varepsilon/k$$

Concentrating on the principal maximum we see that we can adjust the direction of the principle maximum simple by adjusting ε. In a modern phase array radar in fact a dome of antennas are used and the situation is a bit more complicated but certainly a tractable problem with the help of a computer. So these radars have computers adjusting the phases of the various antennas to point the radar beam where desired - which can be much more rapidly scanned than a rotating parabolic antenna for example. We can see that if you increase the number of antennas, then you will get a more narrowly collimated beam.

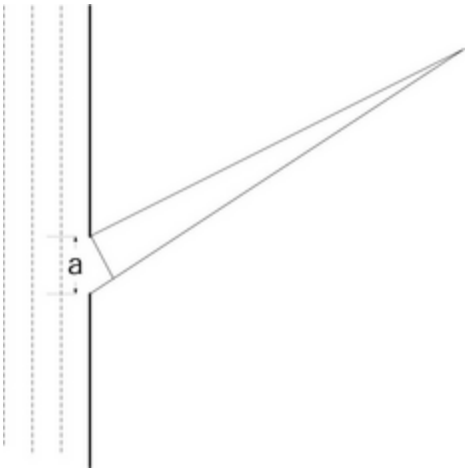


Get a bigger version of this animation.

SingleSlitDiffraction.xhtml

Diffraction

Diffraction is an important characteristic of waves. It can be said to one of the defining characteristics of a wave. It occurs when part of a wavefront is obstructed. The parts of the wavefronts that propagate past the the obstacle interfere and create a diffraction pattern. Diffraction and interference are essentially the same physical process, resulting from the vector addition of fields from different sources. By convention interference refers to only a few sources and diffraction refers to many sources or a continuous source.



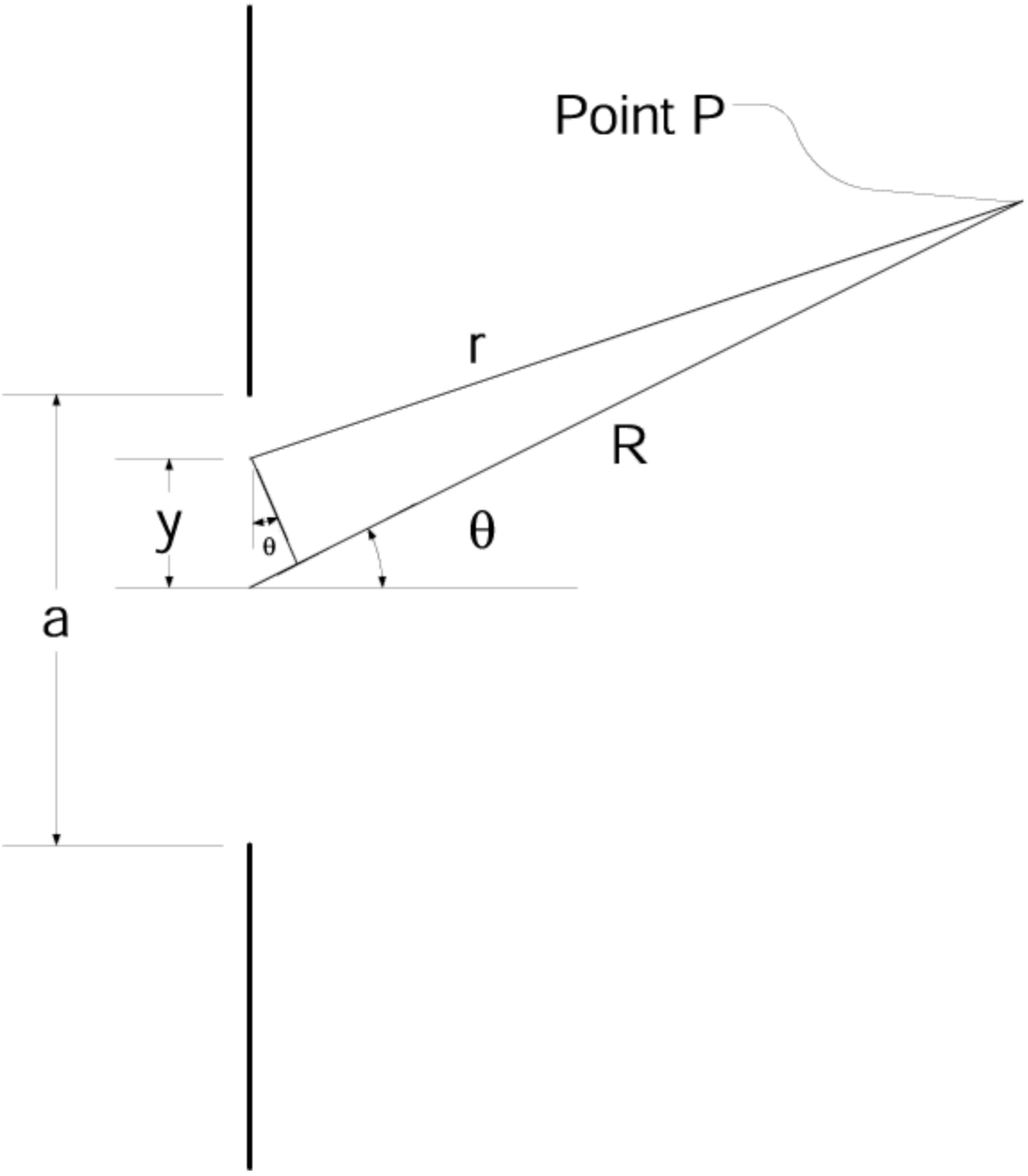
When a plane wave hits an aperture, Huygens principle says that each point in the aperture acts as a source of spherical wavelets. The maximum path length difference of all these sources is between the top and the bottom.

$$\Delta r_{max} = a \sin\theta$$

The waves start out in phase. If $a << \lambda$ then the slit acts as a point source and you get a spherical wave coming out. If $a >> \lambda$ then the aperture simply casts a bright spot the size of the aperture shadow. But if $\lambda \approx a$ then an interference pattern is set up.When the resulting pattern is viewed close to the aperture, the pattern can be very complex, and this is call Fresnel diffraction. As the the pattern is viewed from further and further away, it

eventually stops changing shape and only grows in size. This is Fraunhoffer diffraction.

Single Slit Diffraction



Consider the contribution to the field $\vec{E}$ at a P due to a small element of the slit $\mathbb{d}y$ at y. It is a distance r from P. R is the distance from the center of the slit to P.

lets define ε_L which is the source strength per unit length, which is a constant.

then

$$dE = \frac{\varepsilon_L}{r}\mathbb{d}y e^{i(kr-\omega t)}$$

Now from the drawing

$$\begin{gathered} r^2 = (R - y\sin\theta)^2 + (y\cos\theta)^2 \\ = R^2 + y^2\sin^2\theta - 2Ry\sin\theta + y^2\cos^2\theta \\ = R^2 + y^2 - 2Ry\sin\theta \\ = R^2\left[1 - \tfrac{2y}{R}\sin\theta + \tfrac{y^2}{R^2}\right] \end{gathered}$$

Now assume that $y << R$ (which gives us the Franhaufer condition) and

$$r = R\left[1 - \frac{2y}{R}\sin\theta\right]^{\frac{1}{2}}$$

now expand the square root

$$r = R\left[1 - \frac{y}{R}\sin\theta + \ldots\right]$$

and neglect higher terms so that

$$r = R - y\sin\theta$$

thus

$$dE = \frac{\varepsilon_L}{R}e^{i(k(R-y\sin\theta)-\omega t)}\mathbb{d}y$$

where now we have used R in the denominator since it is much bigger than y

$$dE = \frac{\varepsilon_L}{R} e^{i(kR-\omega t)} e^{-iky\sin\theta} \mathbb{d}y$$

now integrate assuming that θ is a constant over the slit

$$\begin{aligned} E &= \tfrac{\varepsilon_L}{R} e^{i(kR-\omega t)} \int_{-a/2}^{a/2} e^{-iky\sin\theta} \mathbb{d}y \\ &= \tfrac{\varepsilon_L}{R} e^{i(kR-\omega t)} \tfrac{e^{-iky\sin\theta}}{-ik\sin\theta} \Big|_{-a/2}^{a/2} \\ &= \tfrac{\varepsilon_L}{R} e^{i(kR-\omega t)} \tfrac{e^{-i\frac{ka}{2}\sin\theta} - e^{i\frac{ka}{2}\sin\theta}}{-ik\sin\theta} \\ &= \tfrac{\varepsilon_L}{R} e^{i(kR-\omega t)} \tfrac{-2i\sin(\frac{ka}{2}\sin\theta)}{-ik\sin\theta} \\ &= \tfrac{\varepsilon_L}{R} e^{i(kR-\omega t)} \tfrac{2\sin(\frac{ka}{2}\sin\theta)}{k\sin\theta} \\ &= \tfrac{\varepsilon_L a}{R} e^{i(kR-\omega t)} \tfrac{\sin(\frac{ka}{2}\sin\theta)}{\frac{ka}{2}\sin\theta} \end{aligned}$$

now we define

$$\beta = \frac{ka}{2}\sin\theta$$

and see that we can rewrite our expression as

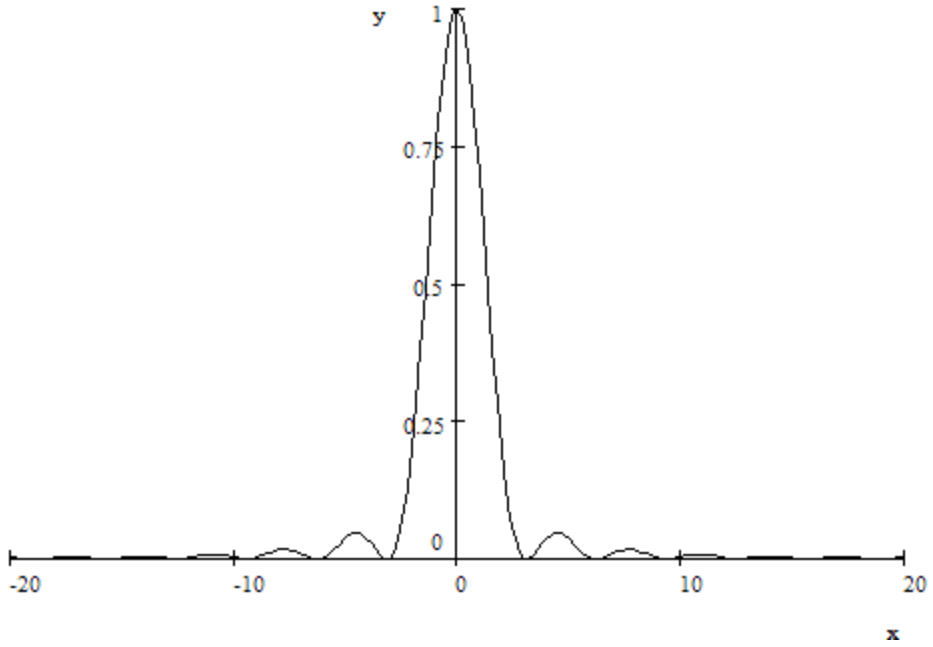
$$E = \frac{\varepsilon_L a}{R} \frac{\sin\beta}{\beta} e^{i(kR-\omega t)}$$

or equivalently

$$E = \frac{\varepsilon_L a}{R} sinc\beta e^{i(kR-\omega t)}$$

The intensity will go like the square of this so

$$I = I_0 sinc^2\beta$$



Plot of $\frac{\sin^2 \beta}{\beta^2}$

The Intensity has a maximum at $\beta = 0$ or $\theta = 0$. there are minima when $\sin \beta = 0$ or

$$\beta = \frac{ka}{2} \sin \theta = n\pi$$

$$\frac{2\pi}{\lambda} \frac{a}{2} \sin \theta = n\pi$$

$$\sin \theta = \frac{n\lambda}{a}$$

in the case of small θ we see that

$$\Delta\theta = \frac{\lambda}{a}$$

is the distance between adjacent minima.

As a becomes large, we see that the minima will merge together. This is consistent with what we said at the beginning, that if

$$a >> \lambda$$

then you just get shadowing but not diffraction.

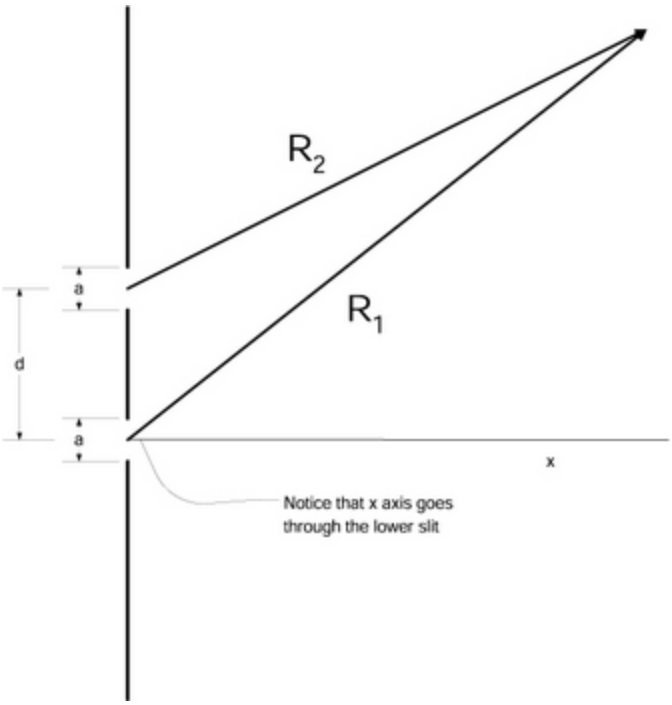
Finding the secondary maxima is more difficult. (Take the derivative of I and then look for zeros.) This can not be done analytically.

Note that wee have been considering only one dimension. If the length of the slit is L then we have only considered the case that $L >> \lambda$ and so diffraction occurs only in the other dimension.

DoubleSlitDiffraction.xhtml

Two Slit Diffraction

Now we consider the case of two slit diffraction.



Notice that the x axis has been drawn through the lower slit. Then the field at the distant point is just the sum of the field from the two slits. Thus we can use our solution to single slit diffraction for each slit and add them together

$$E = \frac{\varepsilon_L a}{R_1} sinc\beta e^{i(kR_1-\omega t)} + \frac{\varepsilon_L a}{R_2} sinc\beta e^{i(kR_2-\omega t)}$$

Now we we will define $R = R_1$ and use $R_2 = R - d\sin\theta$

$$E = \frac{\varepsilon_L a}{R} sinc\beta e^{i(kR-\omega t)} + \frac{\varepsilon_L a}{R - d\sin\theta} sinc\beta e^{i(kR-kd\sin\theta-\omega t)}$$

Now we can ignore the $d \sin\theta$ in the denominator, as it will not have a significant effect on that. However in the exponent, we can not ignore it, since it could significantly affect the phase of the harmonic function. Lets define $\alpha = (kd \sin\theta)/2$ so now we can write:

$$E = \frac{\varepsilon_L a}{R} sinc\beta e^{i(kR-\omega t)} + \frac{\varepsilon_L a}{R} sinc\beta e^{i(kR-2\alpha-\omega t)}$$

and start rearranging:

$$E = \frac{\varepsilon_L a}{R} sinc\beta \left[2\cos\alpha\right] e^{i(kR-\alpha-\omega t)}$$

This is very similar to the case of single slit diffraction except that you now get a factor $2\cos\alpha$ included and a phase shift in the harmonic function.

So we can see immediately the intensity is

$$I = 4I_0 \cos^2\alpha sinc^2\beta$$

recall

$$\alpha = (kd \sin\theta)/2$$

and

$$\beta = (ka \sin\theta)/2$$

If d goes to 0 then expression just becomes the expression for single slit diffraction. If a goes to 0 then the expression just becomes that for Youngs double slit. The double slit diffraction is just the product of these two results. (Hey cool!)

DiffractionGrating.xhtml

Diffraction Grating

Consider the case of N slit diffraction, We have

$$E_1 = \frac{\varepsilon_L a}{R}\frac{\sin\beta}{\beta}e^{i(kR_1-\omega t)}$$

$$E_2 = \frac{\varepsilon_L a}{R}\frac{\sin\beta}{\beta}e^{i(kR_2-\omega t)}$$

$$\vdots$$

$$E_N = \frac{\varepsilon_L a}{R}\frac{\sin\beta}{\beta}e^{i(kR_N-\omega t)}$$

So we can just follow the steps of the two slit case and extend them and get (using $R_N = R-(N-1)d\sin\theta$)

$$\begin{aligned} E &= \textstyle\sum_{n=1}^{N} E_N \\ &= \textstyle\sum_{n=1}^{N} \frac{\varepsilon_L a}{R}\frac{\sin\beta}{\beta}e^{i(kR-2(n-1)\alpha-\omega t)} \\ &= \frac{\varepsilon_L a}{R}\frac{\sin\beta}{\beta}\textstyle\sum_{n=1}^{N} e^{i(kR-2(n-1)\alpha-\omega t)} \\ &= \frac{\varepsilon_L a}{R}\frac{\sin\beta}{\beta}e^{i(kR-\omega t)}\textstyle\sum_{n=1}^{N} e^{-i2(n-1)\alpha} \\ &= \frac{\varepsilon_L a}{R}\frac{\sin\beta}{\beta}e^{i(kR-\omega t)}\textstyle\sum_{j=0}^{N-1} e^{-i2j\alpha} \end{aligned}$$

This is the same geometric series we dealt with before

$$\sum_{n=0}^{N-1} x^n = \frac{1-x^N}{1-x}$$

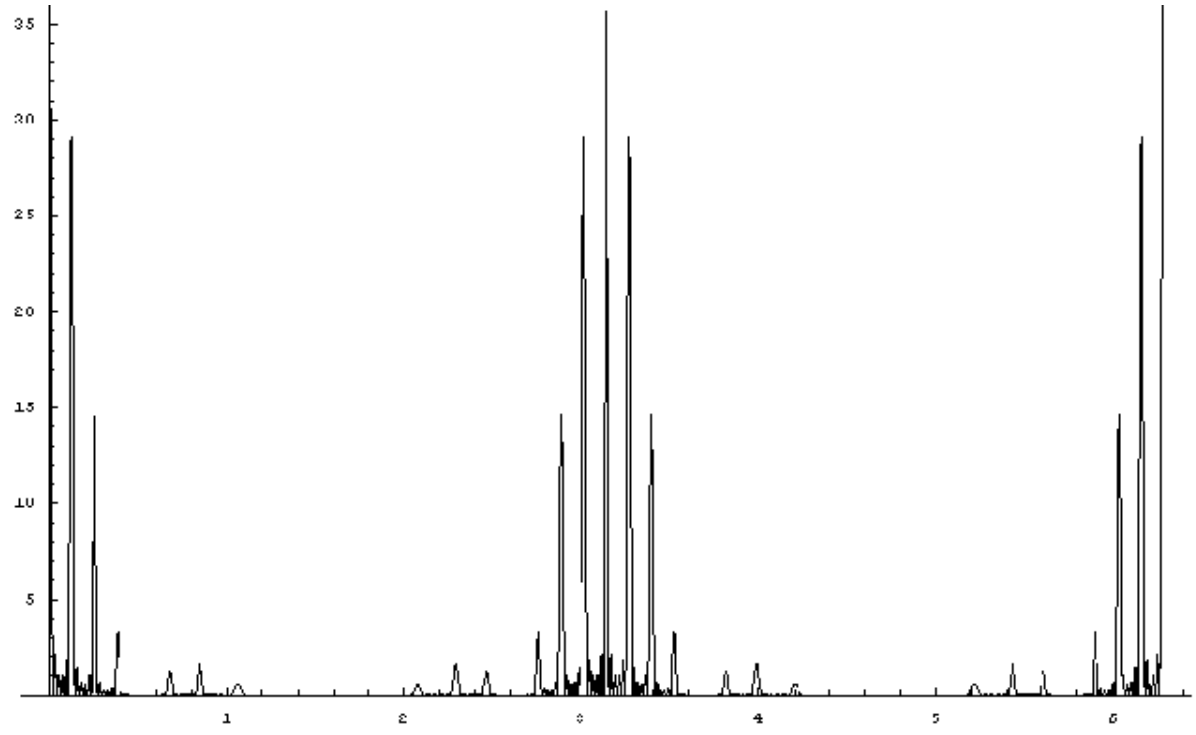
so

$$
\begin{aligned}
E &= \frac{\varepsilon_L a}{R}\,\frac{\sin\beta}{\beta}\,e^{i(kR-\omega t)}\sum_{j=0}^{N-1} e^{-i2j\alpha} \\
&= \frac{\varepsilon_L a}{R}\,\frac{\sin\beta}{\beta}\,e^{i(kR-\omega t)}\,\frac{1-e^{-i2N\alpha}}{1-e^{-i2\alpha}} \\
&= \frac{\varepsilon_L a}{R}\,\frac{\sin\beta}{\beta}\,e^{i(kR-\omega t)}\,\frac{e^{-iN\alpha}}{e^{-i\alpha}}\,\frac{e^{iN\alpha}}{e^{i\alpha}}\,\frac{1-e^{-i2N\alpha}}{1-e^{-i2\alpha}} \\
&= \frac{\varepsilon_L a}{R}\,\frac{\sin\beta}{\beta}\,e^{i(kR-\omega t)}\,\frac{e^{-iN\alpha}}{e^{-i\alpha}}\,\frac{e^{iN\alpha}-e^{-iN\alpha}}{e^{i\alpha}-e^{-i\alpha}} \\
&= \frac{\varepsilon_L a}{R}\,\frac{\sin\beta}{\beta}\,e^{i(kR-(N-1)\alpha-\omega t)}\,\frac{\sin N\alpha}{\sin\alpha}
\end{aligned}
$$

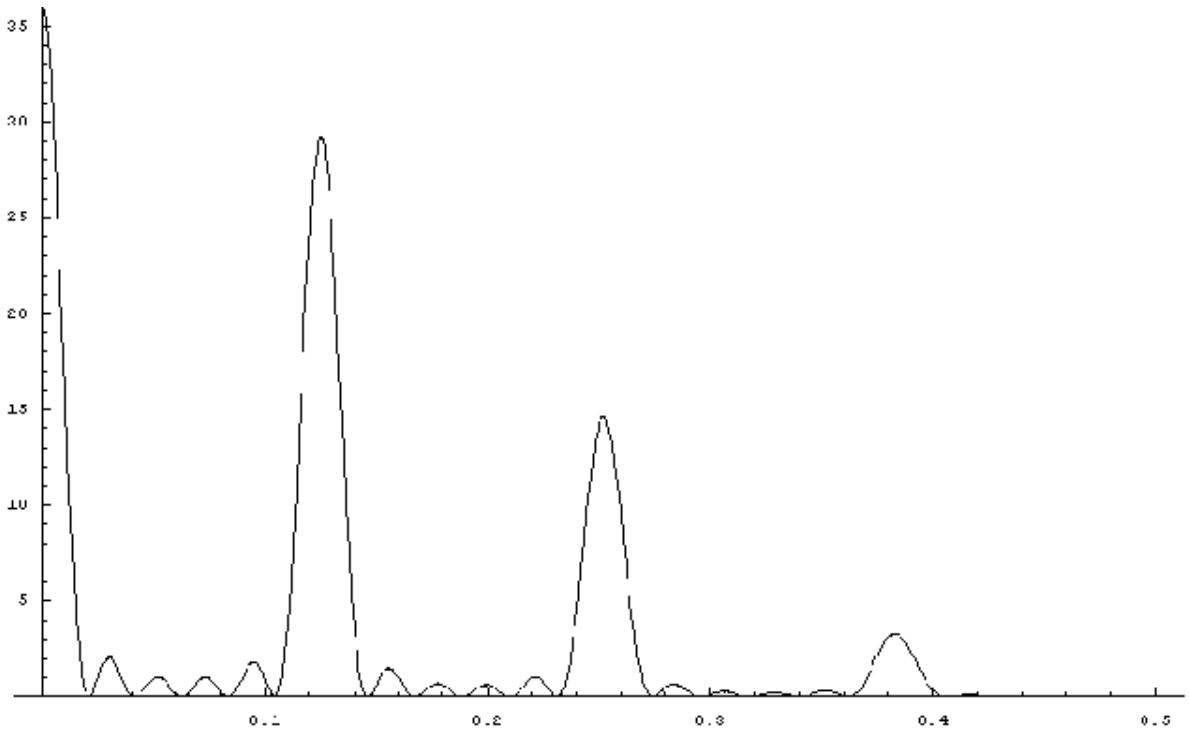
Notice that this just ends up being multisource interference multiplied by single slit diffraction.

Squaring it we see that:

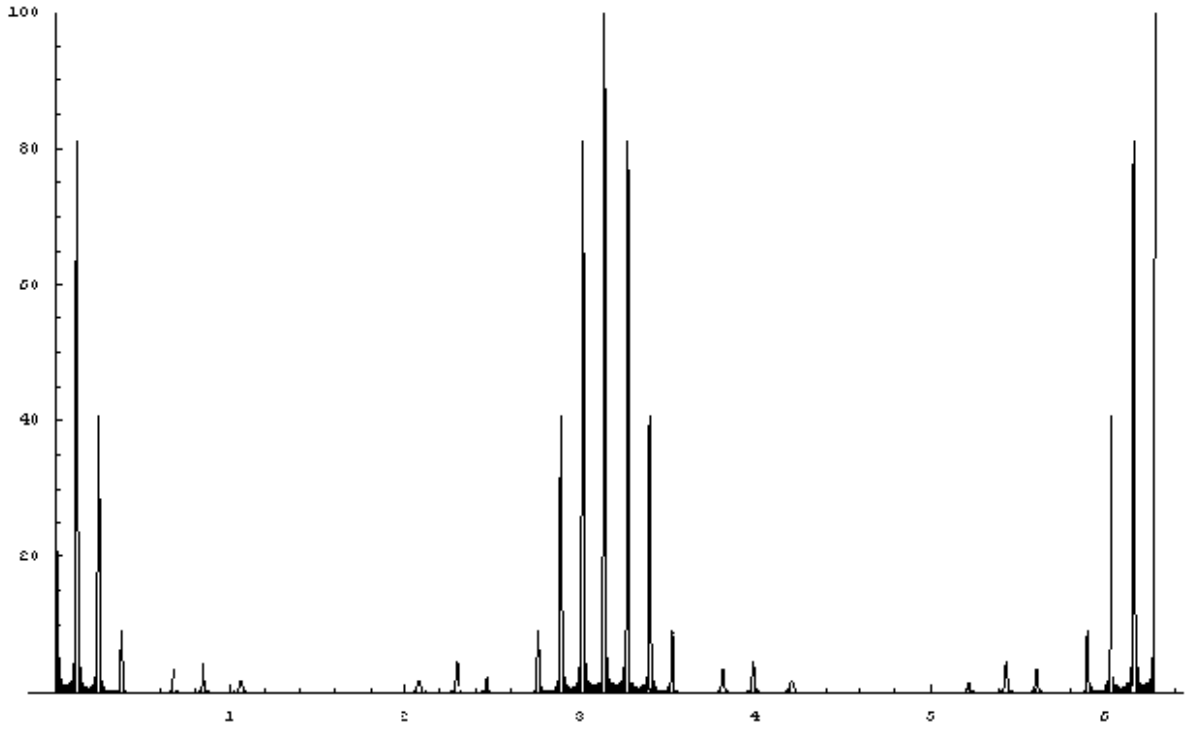
$$
I(\theta) = I_0\,\frac{\sin^2\beta}{\beta^2}\,\frac{\sin^2 N\alpha}{\sin^2\alpha}
$$



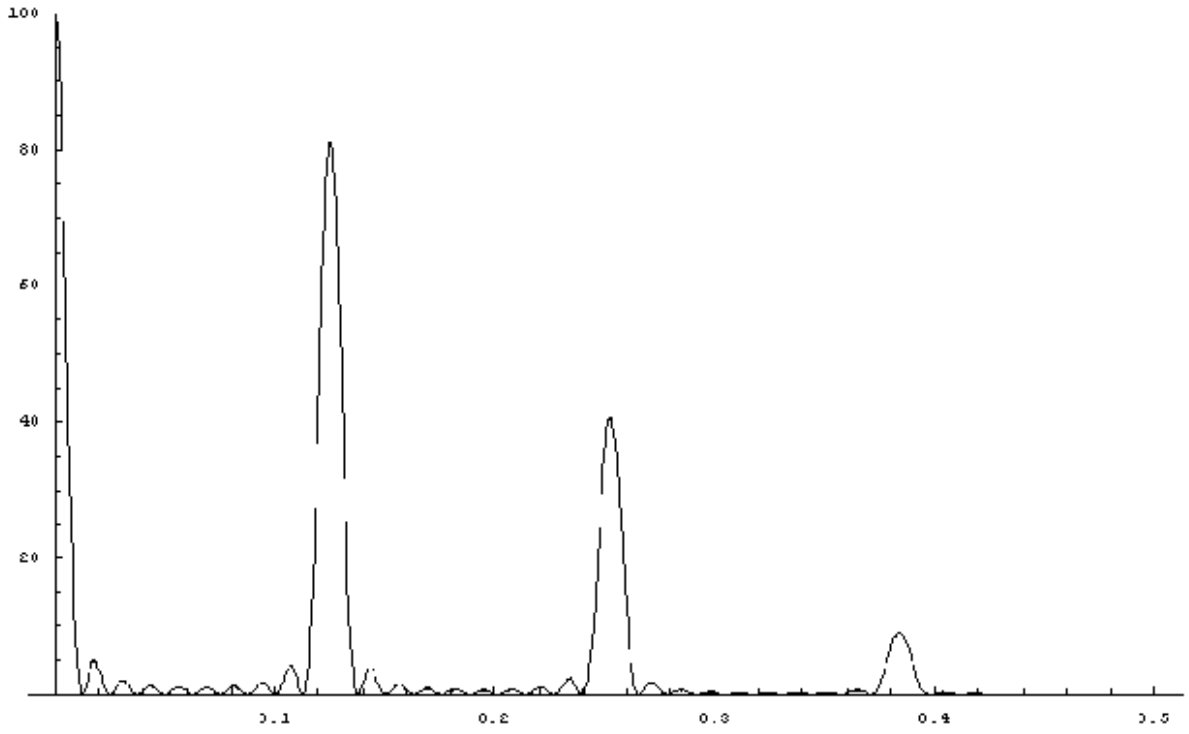
Interference with diffraction for 6 slits with $d = 4a$



Interference with diffraction for 6 slits with $d = 4a$



Interference with diffraction for10 slits with $d = 4a$



Interference with diffraction for10 slits with $d = 4a$

Principal maxima occur when

$$\frac{\sin N\alpha}{\sin \alpha} = N$$

or since $\alpha = kd(\sin\theta)/2$

$$kd\sin\theta = 2n\pi \qquad n = 0, 1, 2, 3$$

or

$$\frac{2\pi}{\lambda} d\sin\theta = 2n\pi$$

or

$$\sin\theta = \frac{n\lambda}{d}$$

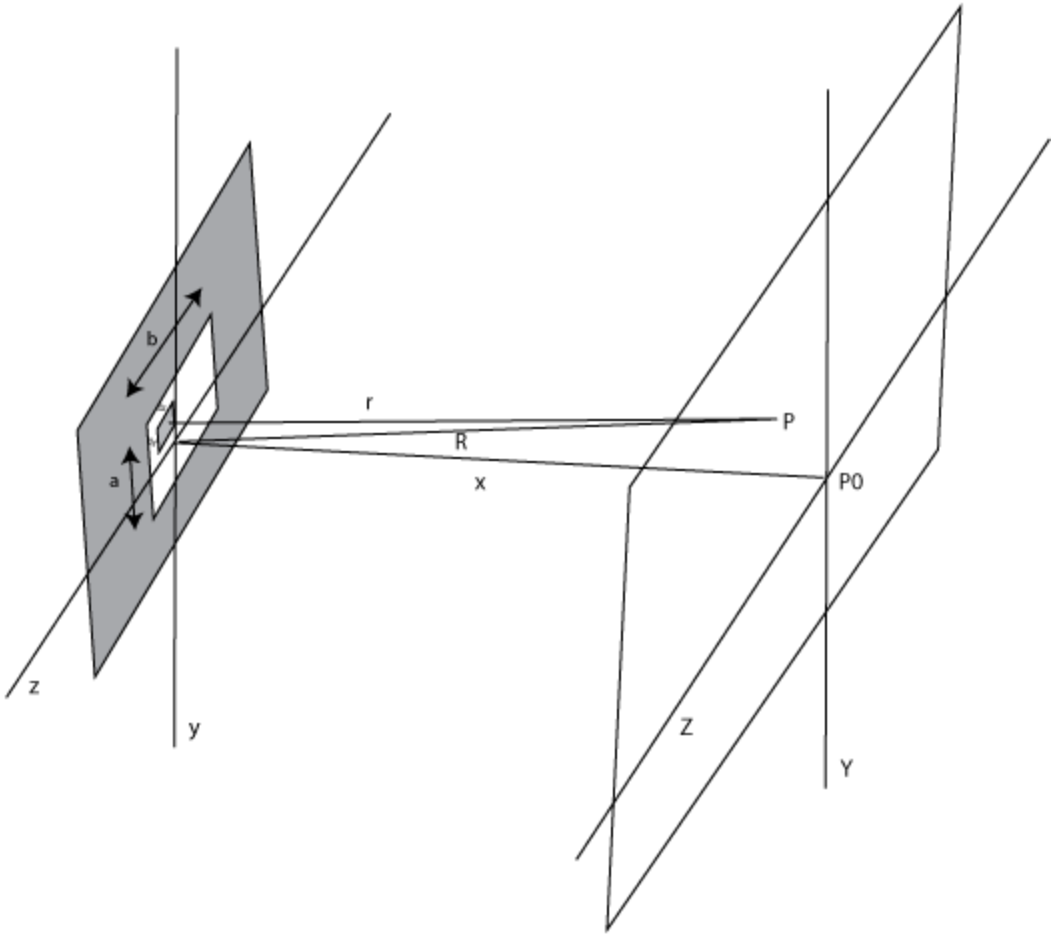
and just like in multisource interference minima occur at

$$\sin\theta = \frac{n\lambda}{Nd} \qquad n = 1, 2, 3\ldots \quad \frac{n}{N} \neq integer$$

A diffraction grating is a repetitive array of diffracting elements such as slits or reflectors. Typically with N very large (100's). Notice how all but the first maximum depend on λ. So you can use a grating for spectroscopy.

Diffraction from a Rectangular Aperture

We consider diffraction from apertures other than a slit. For example consider a rectangular aperture as shown below. If $\vec{\varepsilon}_A$ is the source strength per unit area (assumed to be constant over the entire area in this example) and $dS = \mathbb{d}y\mathbb{d}z$ is an infinitesmal area at a point in the aperture then we have:



$$d\vec{E} = \frac{\vec{\varepsilon}_A}{r} e^{i(kr-\omega t)} \mathbb{d}y\mathbb{d}z$$

We see from the figure that

$$r = \left[x^2 + (Y-y)^2 + (Z-z)^2\right]$$

and that

$$R^2 = x^2 + Y^2 + Z^2.$$

Thus we use

$$x^2 = R^2 - Y^2 - Z^2$$

to write

$$r = \left[R^2 - Y^2 - Z^2 + Y^2 + y^2 - 2yY + Z^2 + z^2 - 2zZ\right]^{1/2}$$

or

$$r = \left[R^2 - 2Yy - 2Zz + y^2 + z^2\right]^{1/2}$$

$$r = R\left[1 - 2Yy/R^2 - 2Zz/R^2 + (y^2 + z^2)/R^2\right]^{1/2}.$$

We are only considering Fraunhofer diffraction so R, Z, Y are much larger than y and z and we can rewrite

$$r \approx R\left[1 - 2(Yy + Zz)/R^2\right]^{1/2}$$

and then finally expanding using the binomial theorem and taking only the most significant terms

$$r \approx R\left[1 - (Yy + Zz)/R^2\right].$$

So we have

$$d\vec{E} = \frac{\vec{\mathcal{E}_A}}{r} e^{i(kr - \omega t)} \mathbb{d}y \mathbb{d}z$$

or using the fact that R is large

$$d\vec{E} = \frac{\vec{\varepsilon_A}}{R} e^{i(kr-\omega t)} \mathbb{d}y \mathbb{d}z$$

which we integrate to get the field

$$\vec{E} = \int_{-b/2}^{+b/2} \int_{-a/2}^{+a/2} \frac{\vec{\varepsilon_A}}{R} e^{i(kr-\omega t)} \mathbb{d}z \mathbb{d}y$$

$$= \frac{\vec{\varepsilon_A} e^{i(kR-\omega t)}}{R} \int_{-a/2}^{+a/2} e^{-ikyY/R} \mathbb{d}y \int_{-b/2}^{+b/2} e^{-ikzZ/R} \mathbb{d}z$$

$$= \frac{\vec{\varepsilon_A} e^{i(kR-\omega t)}}{R} \left[\frac{e^{-ikyY/R}}{-ikY/R} \Big|_{-a/2}^{+a/2} \right] \left[\frac{e^{-ikzZ/R}}{-ikZ/R} \Big|_{-b/2}^{+b/2} \right]$$

$$= \frac{\vec{\varepsilon_A} e^{i(kR-\omega t)}}{R} \left[\frac{e^{-ikaY/2R} - e^{ikaY/2R}}{-ikY/R} \right] \left[\frac{e^{-ikbZ/2R} - e^{ikbZ/2R}}{-ikZ/R} \right]$$

Now lets define $\alpha = kaY/2R$ $\beta = kbZ/2R$ and $\beta = kbZ/2R$ and we see that

$$\vec{E} = \frac{\vec{\varepsilon_A} e^{i(kR-\omega t)}}{R} \left[\frac{e^{i\alpha} - e^{-ik\alpha}}{ikY/R} \right] \left[\frac{e^{i\beta} - e^{-i\beta}}{ikZ/R} \right]$$

or rearranging

$$\vec{E} = \frac{\vec{\varepsilon_A} e^{i(kR-\omega t)}}{R} \left[\frac{e^{i\alpha} - e^{-i\alpha}}{2i\alpha/a} \right] \left[\frac{e^{i\beta} - e^{-i\beta}}{2i\beta/b} \right]$$

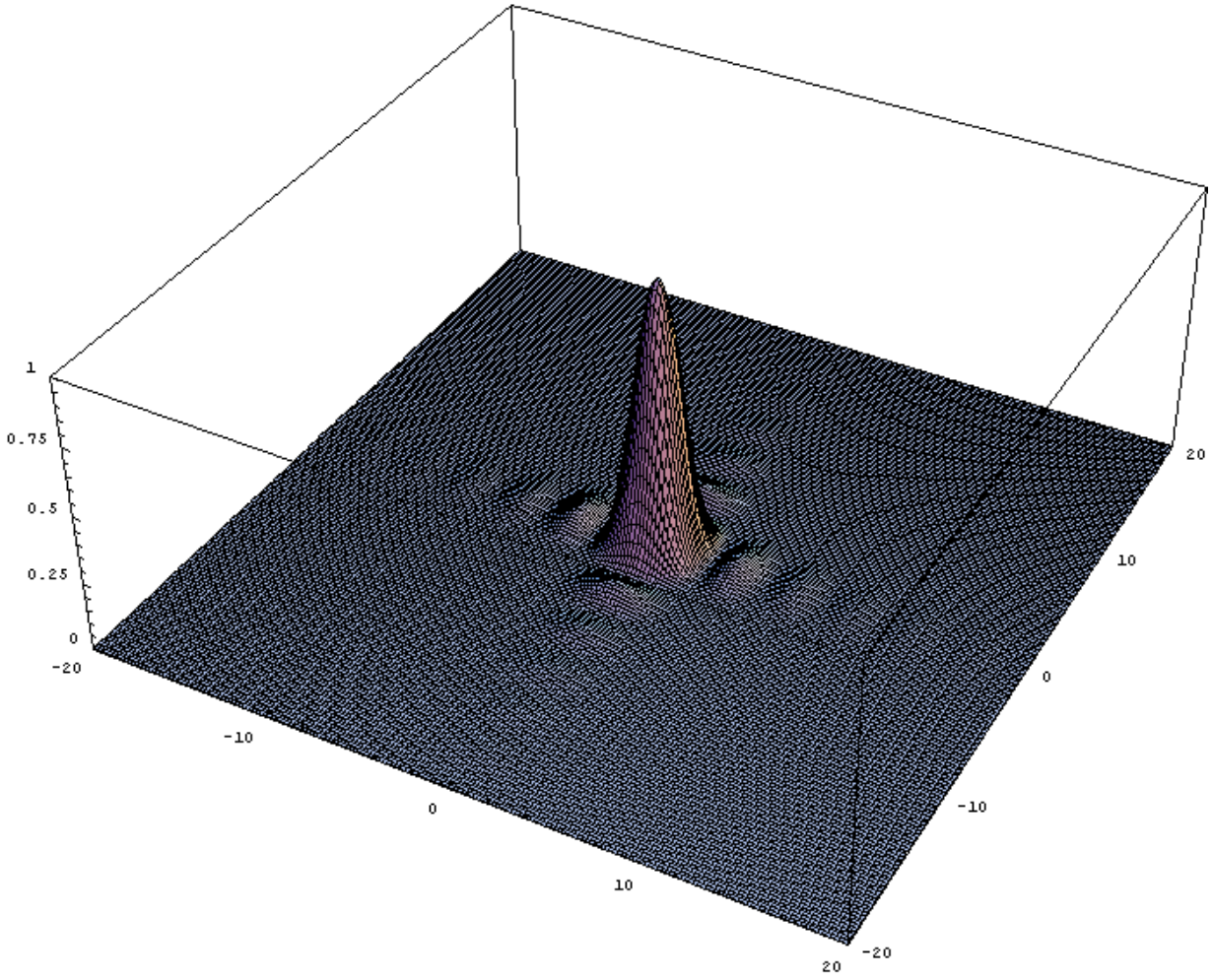
$$\vec{E} = \frac{\vec{\varepsilon_A} e^{i(kR-\omega t)} ab}{R} \left[\frac{e^{i\alpha} - e^{-i\alpha}}{2i\alpha} \right] \left[\frac{e^{i\beta} - e^{-i\beta}}{2i\beta} \right]$$

$$\vec{E} = \frac{\vec{\varepsilon_A} e^{i(kR-\omega t)} ab}{R} \left[\frac{\sin\alpha}{\alpha} \right] \left[\frac{\sin\beta}{\beta} \right]$$

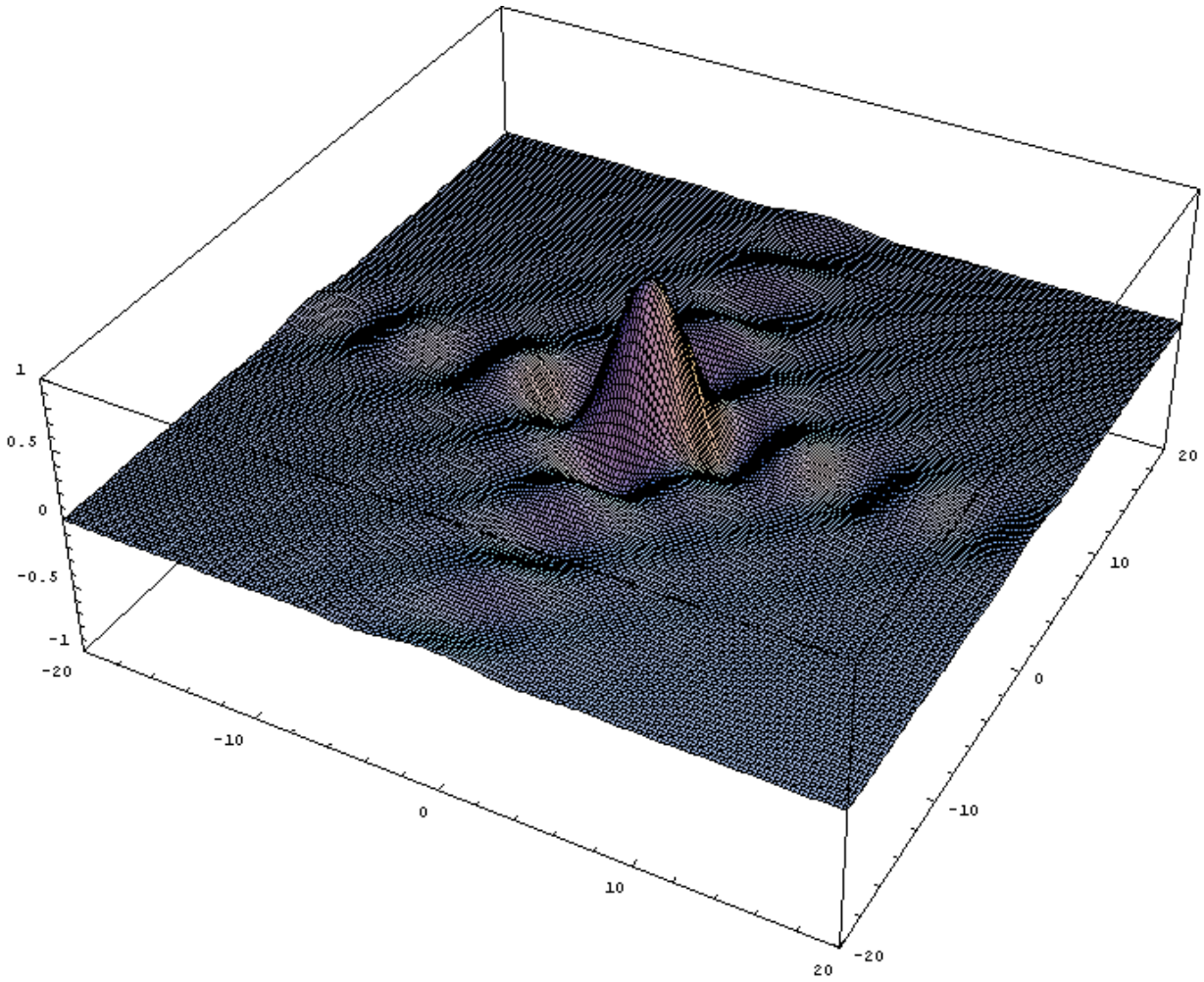
So finally we can write the intensity as

$$I(Y,Z) = I(0,0) \left[\frac{\sin\alpha}{\alpha} \right]^2 \left[\frac{\sin\beta}{\beta} \right]^2$$

Below is a plot of $\left[\frac{\sin\beta}{\beta}\right]^2\left[\frac{\sin\alpha}{\alpha}\right]^2$



Below is a plot of $\left[\frac{\sin\beta}{\beta}\right]\left[\frac{\sin\alpha}{\alpha}\right]$


1
0.5
0
-0.5
-1
-20
-10
0
10
20
-20
-10
0
10
20

Babinet's Principle

Babinet's Principle

Say you have a slit with light passing through it. You will get a diffraction pattern, lets call it $\vec{E}_s$. If you cover the slit with a piece of material that fits just inside the slit, then there is no $\vec{E}$ field in the Fraunhofer limit. The is means that the $\vec{E}$ field of the blocker, lets call it $vecE_b$, must exactly cancel $\vec{E}_s$. The only way this can happen is if

$$\vec{E}_b = -\vec{E}_s$$

. Now if you take the slit away the $\vec{E}$ field of the blocker must still remain and then irradiance must be

$$I_b = (-E_s)^2 = I_s$$

The interference pattern looks the same. You can verify this yourself by taking a strand of your hair and a laser pointer. Shine the laser pointer at a wall and then put a strand of your hair in front of the light beam. The resulting interference pattern is the exact same as one would obtain from a slit with the same width as your hair.

We can use Babinet's Principle to solve complex problems. For example, say you have square aperture with sides of length L. The the diffraction pattern for light passing through it is

$$\vec{E} = \vec{\varepsilon}_A e^{i(kr-\omega t)} \frac{L^2}{R} \left[\frac{\sin \beta_L}{\beta_L} \right] \left[\frac{\sin \alpha_L}{\alpha_L} \right]$$

where (assuming the aperture lies in the $y-z$ plane)

$$\beta_L = kLY/2R$$

$$\alpha = kLZ/2R.$$

Now put an opaque square of length d in the middle of the aperture. Now the resulting $\vec{E}$ field is

$$\vec{E} = \frac{\vec{\varepsilon}_A}{R} e^{i(kr-\omega t)} \left[L^2 \frac{\sin\beta_L}{\beta_L} \frac{\sin\alpha_L}{\alpha_L} - d^2 \frac{\sin\beta_d}{\beta_d} \frac{\sin\alpha_d}{\alpha_d} \right]$$

or

$$I = I_0 \left[L^2 \frac{\sin\beta_L}{\beta_L} \frac{\sin\alpha_L}{\alpha_L} - d^2 \frac{\sin\beta_d}{\beta_d} \frac{\sin\alpha_d}{\alpha_d} \right]^2$$

$$\beta_L = kLY/2R$$

$$\alpha_L = kLZ/2R.$$

$$\beta_d = kdY/2R$$

$$\alpha_d = kdZ/2R.$$

Diffraction from a Circular Aperture

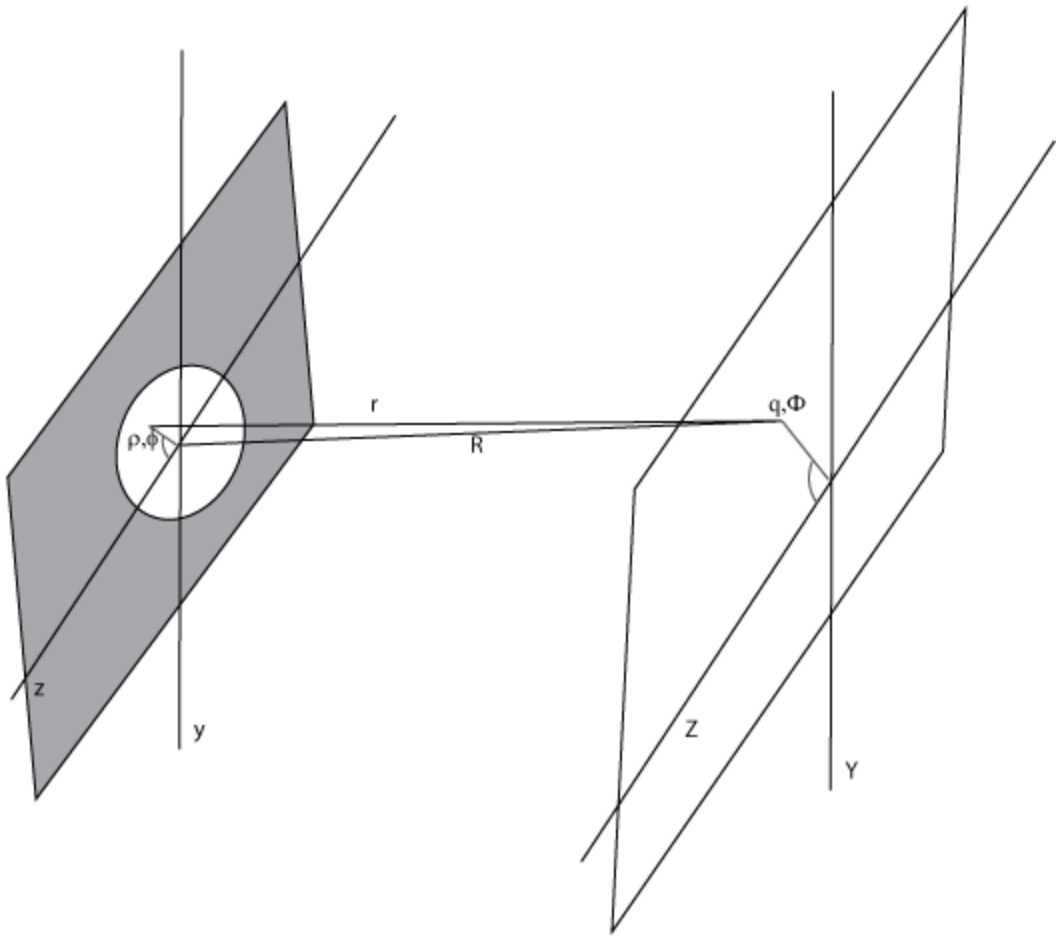
Circular Aperture

The circular aperture is particularly important because it is used a lot in optics. A telescope typically has a circular aperture for example.

We can use the same expression for the E field that we had for the rectangular aperture for any possible aperture, as long as the limits of integration are appropriate. So we can write

$$\vec{E} = \frac{\vec{\mathcal{E}}_A}{R} e^{i(kR-\omega t)} \int\int_{aperture} e^{-iK(Yy+Zz)/R} dy dz$$

For a circular aperture this integration is most easily done with cylindrical coordinates. Look at the figure



Then we have

$$z = \rho \cos \varphi$$

$$y = \rho \sin \varphi$$

$$Z = q \cos \Phi$$

$$Y = q \sin \Phi$$

Then

$$Yy + Zz = \rho q \cos \varphi \cos \Phi + \rho q \sin \varphi \sin \Phi$$

or

$$Yy + Zz = \rho q \cos(\varphi - \Phi)$$

and the integral becomes

$$\vec{E} = \frac{\vec{\varepsilon}_A}{R} e^{i(kR-\omega t)} \int_0^a \int_0^{2\pi} e^{-iK\rho q \cos(\varphi-\Phi)/R} \rho \mathrm{d}\rho \mathrm{d}\varphi$$

In order to do this integral we need to learn a little about Bessel functions.

$$J_0(u) = \frac{1}{2\pi} \int_0^{2\pi} e^{iu\cos v} \mathrm{d}v$$

Is the definition of a Bessel function of the first kind order 0.

$$J_m(u) = \frac{1}{2\pi} \int_0^{2\pi} e^{i(mv + u\cos v)} \mathrm{d}v$$

Is the definition of a Bessel function of the first kind order m.

They have a number of interesting properties such as the recurrence relations

$$\frac{\mathrm{d}}{\mathrm{d}u}[u^m J_m(u))] = u^m J_{m-1}(u)$$

so that for example when $m = 1$

$$\int_0^u u' J_0(u') \mathrm{d}u' = u J_1(u).$$

In order to numerically calculate the value of a Bessel function one uses the expansion

$$J_m(x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!(m+s)!} \left(\frac{x}{2}\right)^{m+2s}.$$

Now we want to evaluate the integral

$$\vec{E} = \frac{\vec{\varepsilon}_A}{R} e^{i(kR-\omega t)} \int_0^a \int_0^{2\pi} e^{-iK\rho q \cos(\varphi - \Phi)/R} \rho d\rho d\varphi$$

which we can do at any value of Φ since the problem is symmetric about Φ. So we can simplify things greatly if we do the integral at $\Phi = 0$

$$\vec{E} = \frac{\vec{\varepsilon}_A}{R} e^{i(kR-\omega t)} \int_0^a \int_0^{2\pi} e^{-iK\rho q \cos(\varphi)/R} \rho d\rho d\varphi$$

which becomes

$$\vec{E} = \frac{\vec{\varepsilon}_A}{R} e^{i(kR-\omega t)} 2\pi \int_0^a J_0(-K\rho q/R) \rho d\rho$$

Now J_0 is an even function so we can drop the minus sign and rewrite the expression as

$$\vec{E} = \frac{\vec{\varepsilon}_A}{R} e^{i(kR-\omega t)} 2\pi \int_0^a J_0(K\rho q/R) \rho d\rho$$

To do this integral we change variables

$$w = k\rho q/R$$

$$\rho = \frac{wR}{kq}$$

$$d\rho = \frac{R}{kq} dw$$

so that

$$\begin{aligned}\int_0^a J_0(K\rho q/R)\rho d\rho &= \int_0^{kaq/R} \left(\tfrac{R}{kq}\right)^2 J_0(w) w dw \\ &= \left(\tfrac{R}{kq}\right)^2 \left(\tfrac{kaq}{R}\right) J_1(kaq/R) \\ &= a^2 \left(\tfrac{R}{kaq}\right) J_1(kaq/R) \\ &= a^2 \tfrac{J_1(kaq/R)}{kaq/R}\end{aligned}$$

So finally we have the result

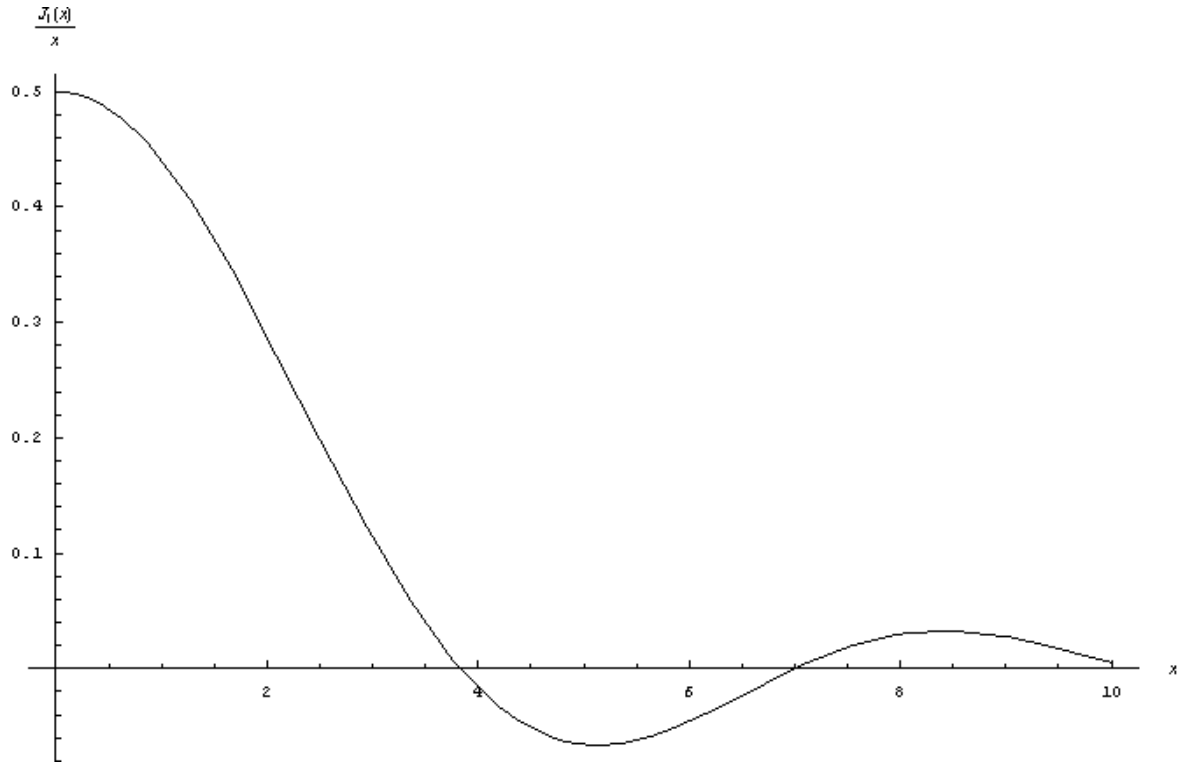
$$\vec{E} = \vec{\varepsilon}_A \frac{e^{i(kR-\omega t)}}{R} 2\pi a^2 \frac{J_1(kaq/R)}{kaq/R}$$

Or recognizing that πa^2 is the area of the aperture A and squaring to get the intensity we write

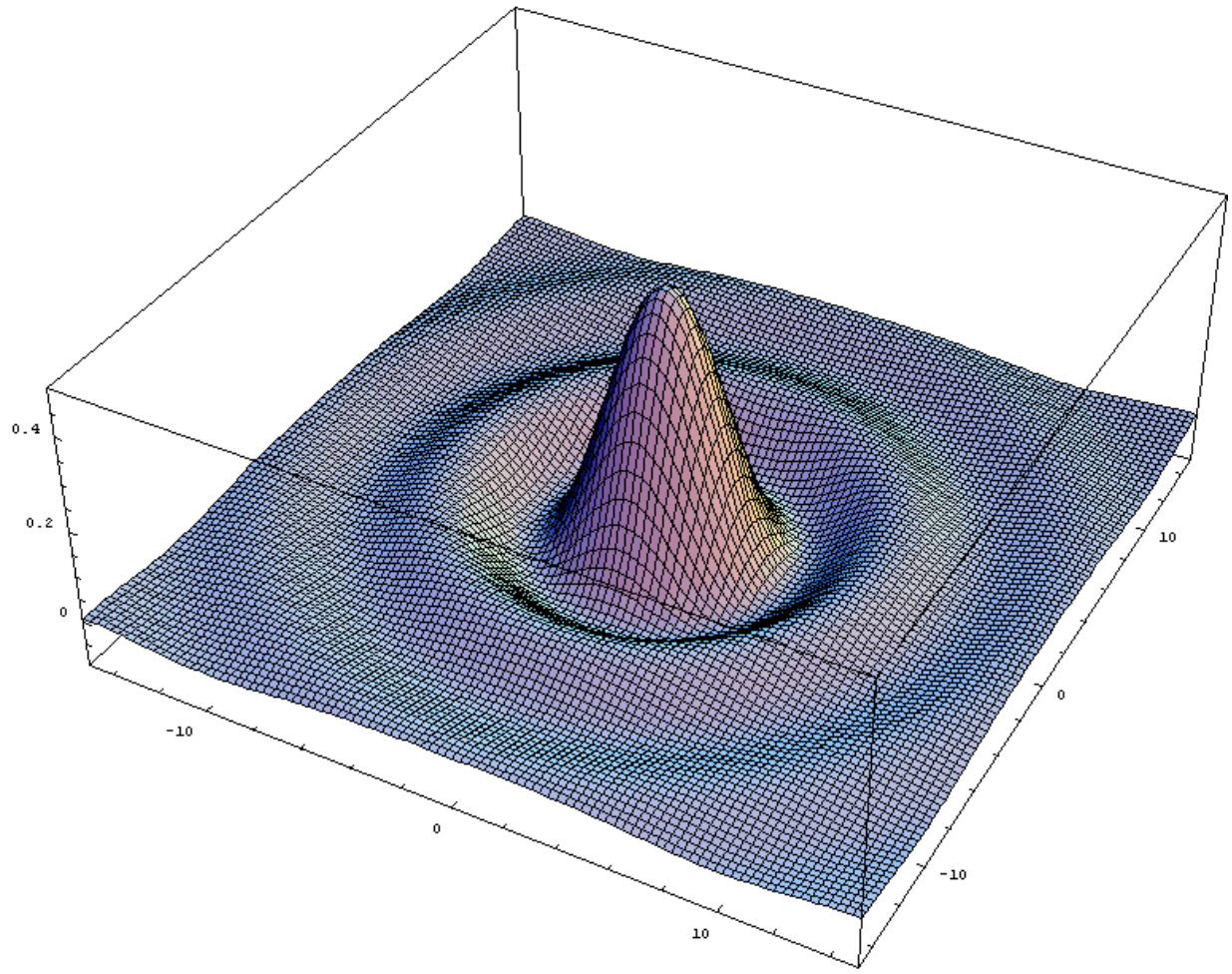
$$I = I_0 \left[\frac{2J_1(kaq/R)}{kaq/R} \right]^2$$

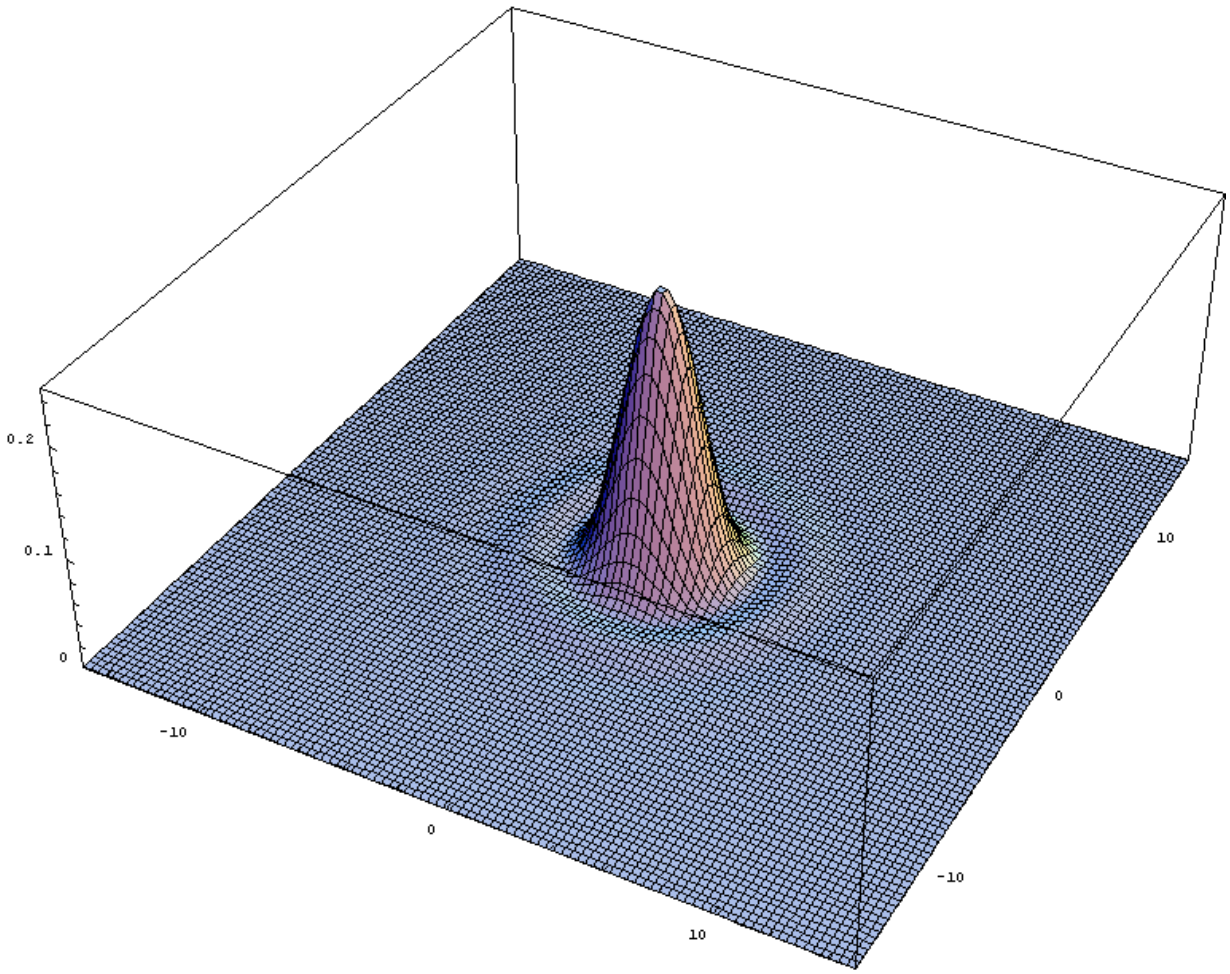
If you want to write this in terms of the angle θ then one uses the fact that $q/R = \sin\theta$

$$I(\theta) = I(0) \left[\frac{2J_1(ka\sin\theta)}{ka\sin\theta} \right]^2$$



Above is a plot of the function $J_1(x)/x$. Notice how it peaks at $1/2$ which is why there is the factor of two in the expression for the irradiance. Below is a 3D plot of the same thing (ie. $J_1(r)/r$). Notice the rings.


0.4
0.2
0
-10
0
10
10
0
-10



Above is a plot of $\left(J_1(r)/r\right)^2$ which corresponds to the irradiance one sees. The central peak out to the first ring of zero is called the Airy disk. This occurs at $J_1(r)/r = 0$ which can be numerically evaluated to give $r = 3.83$ for the first ring.

For our circular aperture above this means the first zero occurs at

$$kaq_1/R = 3.83$$

or

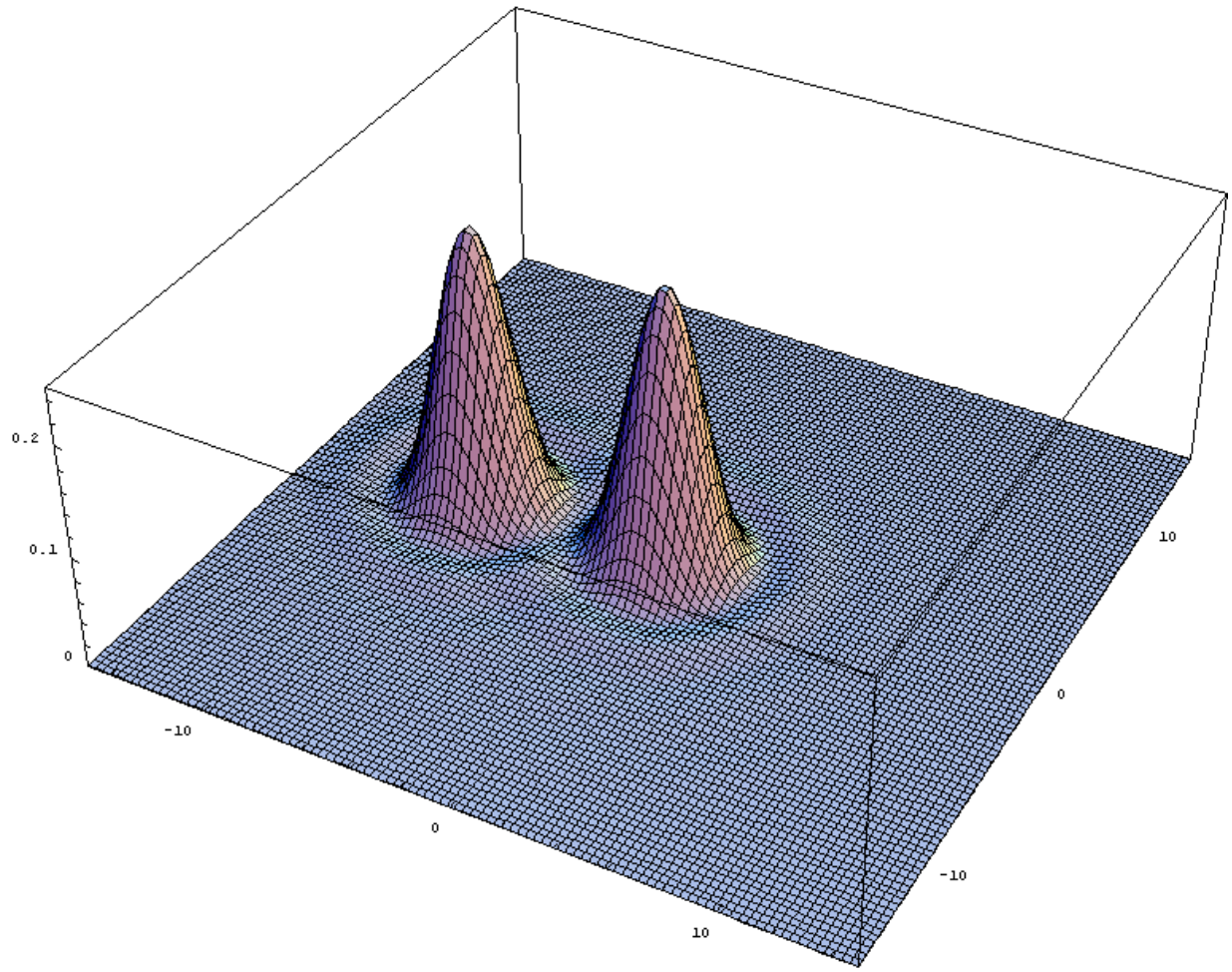
$$\frac{2\pi}{\lambda}\frac{aq_1}{R} = 3.83$$

$$q_1 = \frac{1.22 R \lambda}{2a}$$

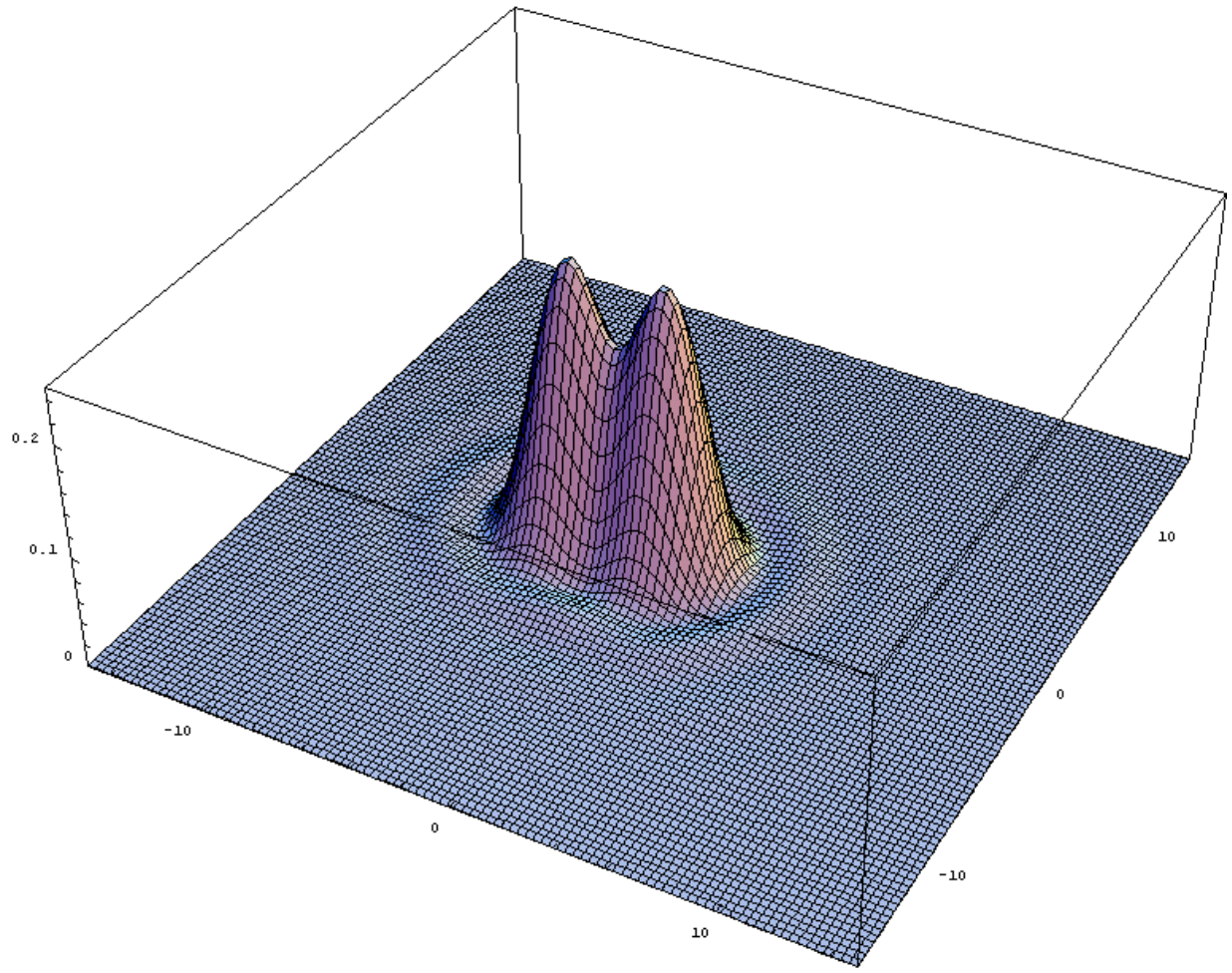
In our case a is the radius of the aperture and we can rewrite the expression using the diameter $D = 2a$

$$q_1 = 1.22 \lambda R / D$$

Light passing through any circular aperture is going to be diffracted in this manner and this sets the limit of resolution on an optical device such as a telescope. Say one is trying resolve two sources, we can say the limit of resolution is when the central spot of one Airy disk is on the zero of the other Airy disk. This is known as the Raleigh critereon. While it is possible to define other crtieria, this is the most commenly used. See for example the plots below



In the above plot, the two sources can clearly be resolved. In the plot below, the two sources are going to be difficult to resolve.



So we say that the limit of our resolution occurs when the distance Δq between two sources is

$$\Delta q = 1.22 R\lambda / D$$

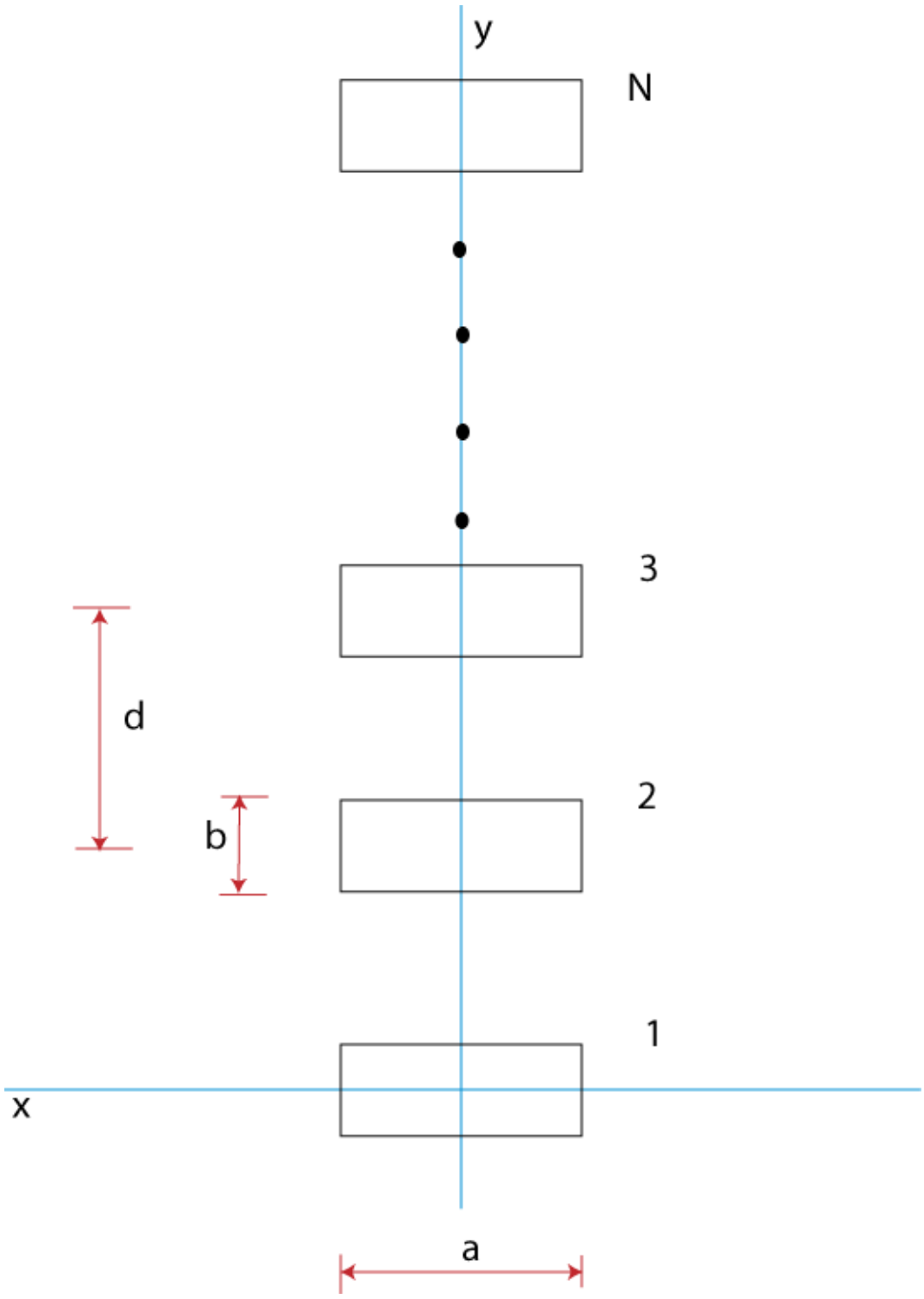
or in the small angle limit $\Delta\theta = \Delta q / R$

$$\Delta\theta = 1.22\lambda / D$$

Diffraction from an Array of Apertures

An array of rectangular apertures

Say we have an array of rectangular apertures sitting in the $x-y$ plane and light hits this aperture traveling in the positive z direction. There are N apertures arranged vertically (in the y direction). Each aperture has a width in the x direction of a and a height in the y direction of b. For convenience, the apertures are aligned with their centers at $x=0$. The apertures are equally spaced by a distance d.



The electric field at some point P away from the array is

$$E = \sum_{n_y=1}^{N} E_{n_y}(a,b)$$

where $E_{n_y}(a,b)$ is the field from the n_y slit at that point. The y position of the center of the aperture is $(n_y - 1)d$ so we write $y = (n_y - 1)d + y'$ and (x, y') is the position of a point in the aperture with respect to the center of the aperture. We can write

$$E_{n_y}(a,b) = \int_{-b/2}^{b/2} \mathrm{d}y' \int_{-a/2}^{a/2} \mathrm{d}x \frac{\mathcal{E}_A}{R} e^{-i(kr_{n_y} - \omega t)}$$

If the point of observation is (x_P, y_P, z_P) then

$$r_{n_y} = \left[(x_P - x)^2 + (y_P - y)^2 + (z_P - z)^2\right]^{1/2}$$

but we take z to be zero at the aperture so

$$\begin{aligned} r_{n_y} &= \left[(x_P - x)^2 + (y_P - y)^2 + z_P^2\right]^{1/2} \\ &= \left[x_P^2 - 2xx_P + x^2 + y_P^2 - 2yy_P + y^2 + z_P^2\right]^{1/2} \\ &= R\left[1 - 2xx_P/R^2 - 2yy_P/R^2 + (x^2 + y^2)/R^2\right]^{1/2} \end{aligned}$$

where $R^2 = x_P^2 + y_P^2 + z_P^2$, the distance from the origin. In the far field approximation $(x^2 + y^2)/R^2 = 0$ and we can write:

$$r_{n_y} \approx R\left[1 - 2xx_P/R^2 - 2yy_P/R^2\right]^{1/2}.$$

We use the first two terms in the binomial expansion and get

$$\begin{aligned} r_{n_y} &\approx R[1 - xx_P/R^2 - yy_P/R^2] \\ &= R - xx_P/R - yy_P/R \\ &= R - xx_P/R - [(n_y - 1)d + y']y_P/R \\ &= R - xx_P/R - (n_y - 1)\mathrm{d}y_P/R - y'y_P/R \end{aligned}$$

so now we have

$$E_{n_y}(a,b) = \int_{-b/2}^{b/2} \mathrm{d}y' \int_{-a/2}^{a/2} \mathrm{d}x \frac{\mathcal{E}_A}{R} e^{-i(k(R - x x_P/R - (n_y-1)\mathrm{d}y_P/R - y' y_P/R) - \omega t)}.$$

We rearrange to get

$$E_{n_y}(a,b) = \frac{\mathcal{E}_A}{R} e^{-i(kR-\omega t)} e^{ik(n_y-1)\mathrm{d}y_P/R} \int_{-b/2}^{b/2} \mathrm{d}y' e^{iky' y_P/R} \int_{-a/2}^{a/2} \mathrm{d}x e^{ikx x_P/R}.$$

We define $kx_P/R = k_x$ and $ky_P/R = k_y$ so that

$$E_{n_y}(a,b) = \frac{\mathcal{E}_A}{R} e^{-i(kR-\omega t)} e^{ik_y(n_y-1)d} \int_{-b/2}^{b/2} \mathrm{d}y' e^{ik_y y'} \int_{-a/2}^{a/2} \mathrm{d}x' e^{ik_x x'}$$

$$E = \frac{\mathcal{E}_A}{R} e^{-i(kR-\omega t)} \sum_{n_y=1}^{N} e^{ik_y(n_y-1)d} \int_{-b/2}^{b/2} \mathrm{d}y' e^{ik_y y'} \int_{-a/2}^{a/2} \mathrm{d}x e^{ik_x x}.$$

We see that each piece of this is something we did before

$$E = \frac{\mathcal{E}_A}{R} e^{-i(kR-\omega t)} e^{ik_y(N-1)d/2} \frac{\sin\left(Nk_y d/2\right)}{\sin\left(k_y d/2\right)} b sinc(k_y b/2) a sinc(k_x a/2)$$

or if we define

$$kR_c = kR - (N-1)dk_y/2$$

$$E = \frac{\mathcal{E}_A ab}{R} e^{-i(kR_c-\omega t)} \frac{\sin\left(Nk_y d/2\right)}{\sin\left(k_y d/2\right)} sinc(k_y b/2) sinc(k_x a/2)$$

Fourier Optics

Fourier Transforms

The Fourier Transform can be used to represent any well behaved function $f(x)$.

$$f(x) = \frac{1}{\pi}\int_0^\infty [A(k)\cos(kx) + B(k)sin(kx)]dk$$

where

$$A(k) = \int_{-\infty}^{\infty} f(x)\cos(kx)\mathbb{d}x$$

$$B(k) = \int_{-\infty}^{\infty} f(x)\sin(kx)\mathbb{d}x$$

I can now substitute for A and B in the original expression and write:

$$f(x) = \frac{1}{\pi}\int_0^\infty \cos kx \int_{-\infty}^{\infty} f(x')\cos kx'\mathbb{d}x'\mathbb{d}k + \\ + \frac{1}{\pi}\int_0^\infty \sin kx \int_{-\infty}^{\infty} f(x')\sin kx'\mathbb{d}x'\mathbb{d}k$$

and then use

$$\cos(x' - x) = coskx\cos kx' + \sin kx \sin kx'$$

$$f(x) = \frac{1}{\pi}\int_0^\infty \int_{-\infty}^{\infty} f(x')\cos(k[x' - x])\mathbb{d}x'\mathbb{d}k$$

Since the inner integral is an even function we can write

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^\infty \int_{-\infty}^{\infty} f(x')\cos(k[x' - x])\mathbb{d}x'\mathbb{d}k$$

Now consider the fact that

$$\frac{i}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x')\sin(k[x'-x])\mathrm{d}x'\mathrm{d}k = 0$$

because $\sin$ is an odd function, ie.

$$\int_{-\infty}^{\infty}\sin(k[x'-x])\mathrm{d}k = 0$$

So we could have written

$$\begin{aligned} f(x) = \tfrac{1}{2\pi}\textstyle\int_{\infty}^{\infty}\int_{-\infty}^{\infty} f(x')\cos(k[x'-x])\mathrm{d}x'\mathrm{d}k + \\ + \tfrac{i}{2\pi}\textstyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x')\sin(k[x'-x])\mathrm{d}x'\mathrm{d}k \end{aligned}$$

or

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x')e^{ik(x'-x)}\mathrm{d}x'\mathrm{d}k$$

or

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(k)e^{-ikx}\mathrm{d}k$$

where

$$g(k) = \int_{-\infty}^{\infty} f(x)e^{ikx}\mathrm{d}x$$

is the Fourier transform of $f(x)$.

Symbolically we write

$$g(k) = F\{f(x)\}$$

$$f(x) = F^{-1}\{g(k)\} = F^{-1}\{F\{f(x)\}\}$$

Now these concepts are easily extended to two dimensions:

$$f(x,y) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(k_x,k_y)e^{-i(xk_x+yk_y)}\mathbb{d}k_x\mathbb{d}k_y$$

where

$$g(k_x,k_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{i(xk_x+yk_y)}\mathbb{d}x\mathbb{d}y.$$

This tells us is that any nonperiodic function of two variables $f(x,y)$ can be synthesized from a distribution of plane waves each with amplitude $g(k_x,k_y)$.

Lets consider Fraunhofer diffraction through an aperture again. For example consider a rectangular aperture as show in the figure. If $\vec{\mathcal{E}}_A$ is the source strength per unit area (assumed to be constant over the entire area in this example) and $dS = \mathbb{d}x\mathbb{d}z$ is an infinitesmal area at a point in the aperture then we have:

[missing_resource: graphics/RectangularAperture.png]

We can define a source strength per unit area

$$d\vec{E} = \frac{\vec{\mathcal{E}}_A}{R}e^{i(\omega t - kr)}$$

Notice that I flipped the sign in the exponential from what I used in the earlier lectures on diffraction. This does not change the physics content of what we are doing in any way, however it allows our notation to follow standard convention.

$$\vec{E} = \int\int_{Aperture}\frac{\vec{\mathcal{E}}_A}{R}e^{i(\omega t - kr)}\mathbb{d}z\mathbb{d}y$$
$$= \frac{\vec{\mathcal{E}}_A e^{i(\omega t - kR)}}{R}\int\int_{Aperture} e^{ikyY/R}\mathbb{d}y e^{ikzZ/R}\mathbb{d}z$$

If we define $k_y = kY/R$ and $k_z = Z/R$ and we see that

$$\overrightarrow{E} = \frac{\vec{\varepsilon_A} e^{i(\omega t - kR)}}{R} \int \int_{Aperture} e^{iyk_y} e^{izk_z} \mathbb{d}y \mathbb{d}z$$

$$\overrightarrow{E} = \frac{\vec{\varepsilon_A} e^{i(\omega t - kR)}}{R} \int \int_{Aperture} e^{i(yk_y + zk_z)} \mathbb{d}y \mathbb{d}z$$

That is, it is equal to the Fourier transform. In fact one can define an "Aperture Function" $A(y, z)$.Such that

$$\overrightarrow{E} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(y, z) e^{i(yk_y + zk_z)} \mathbb{d}y \mathbb{d}z$$

For a rectangular aperture $A(y, z) = \frac{\vec{\varepsilon_A} e^{i(\omega t - kR)}}{R}$ inside the aperture and zero outside it. The aperture function can be much more complex (literally) allowing for changes in source strength and phase through the aperture. The resulting E field is the Fourier transform of the aperture function.

Dirac Delta Function

The above allows us to relate the Fourier transform of an Aperture and the resulting E field from diffraction through that aperture. To extend this to an array of apertures, requires that one introduce a new concept, the Dirac delta function.

The fundamental definition of the Dirac delta function is

$$\delta(x) = 0 \qquad if \quad x \neq 0$$
$$f(0) = \int_{-\infty}^{\infty} f(x)\delta(x)\mathbb{d}x$$

As a special case if $f(x) = 1$

$$\int_{-\infty}^{\infty} \delta(x)\mathbb{d}x = 1$$

This function has some important properties:

$$f(x') = \int_{-\infty}^{\infty} f(x)\delta(x - x')\mathbb{d}x$$

which follows direction from the definition ie. define a new coordinate

$$a = x - x'$$

, then

$$x = a + x'$$

and

$$da = \mathbb{d}x.$$

Then

$$\int_{-\infty}^{\infty} f(x)\delta(x - x')\mathbb{d}x = \int_{-\infty}^{\infty} f(a + x')\delta(a)\mathbb{d}a$$
$$= f(x')$$

We also note that $\delta(x) = \delta(-x)$.

It gets even more interesting. Recall

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x') e^{ik(x'-x)} \mathrm{d}x' \mathrm{d}k$$

which we rearrange suggestively

$$\begin{aligned} f(x) &= \tfrac{1}{2\pi} \textstyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x') e^{ik(x'-x)} \mathrm{d}x' \mathrm{d}k \\ &= \textstyle\int_{-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x'-x)} \mathrm{d}k \right] f(x') \mathrm{d}x' \end{aligned}$$

we also have

$$f(x) = \int_{-\infty}^{\infty} f(x') \delta(x' - x) \mathrm{d}x'$$

so we must have

$$\delta(x' - x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x'-x)} \mathrm{d}k$$

or

$$\delta(x) = \delta(-x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx} \mathrm{d}k$$

That is the Dirac delta function is the inverse Fourier transform of 1. This is a very useful property that allows us to do problems like Young's double slit. Consider the aperture function:

$$A = E_0(\delta(y - a/2) + \delta(y + a/2))$$

where we have represented the slits by Dirac delta functions. Then we obtain

$$
\begin{aligned}
E &= \int_{-\infty}^{\infty} E_0 \delta(y - a/2) e^{iyk_y} + \int_{-\infty}^{\infty} E_0 \delta(y + a/2) e^{iyk_y} \\
&= E_0 (e^{iak_y/2} + e^{-iak_y/2}) \\
&= E_0 2 \cos\left(k_y a/2\right) \\
&= E_0 2 \cos\left(\tfrac{ka \sin\theta}{2}\right)
\end{aligned}
$$

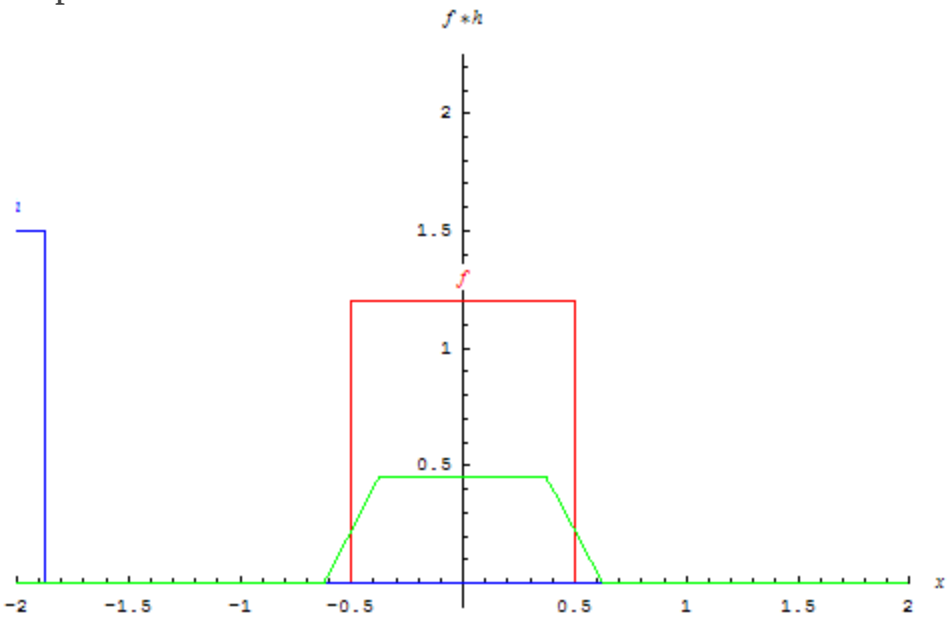
which is exactly what we obtained before. Well that was a lot easier than what we did earlier in the course!

The Convolution Theorem and Diffraction

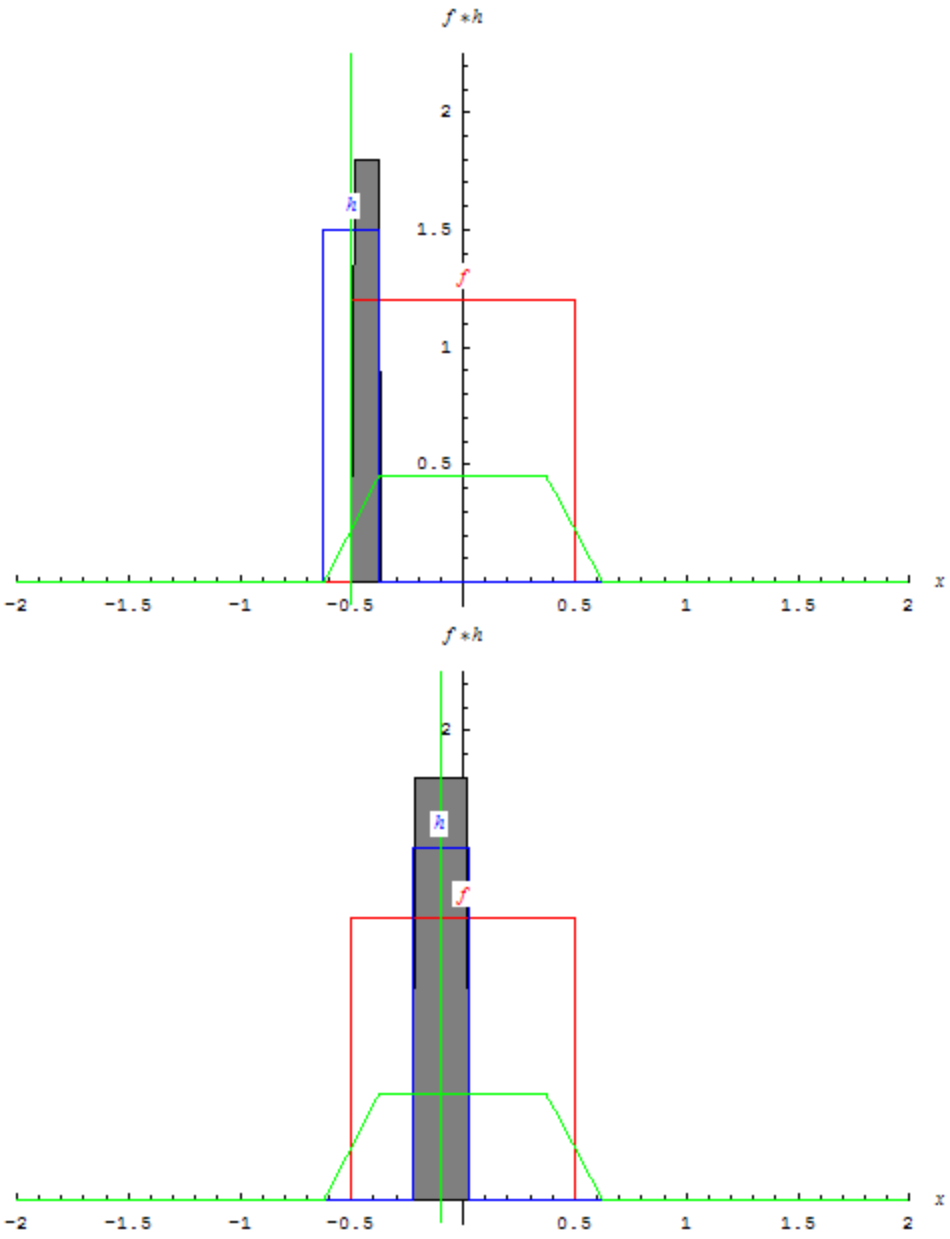
To handle more complex cases of diffraction using Fourier transforms we need to know the convolution theorem. Say $g(x)$is the convolution of two other functions f and h. Then

$$g(x) = f \otimes h = \int_{-\infty}^{\infty} f(x')h(x - x')dx'$$

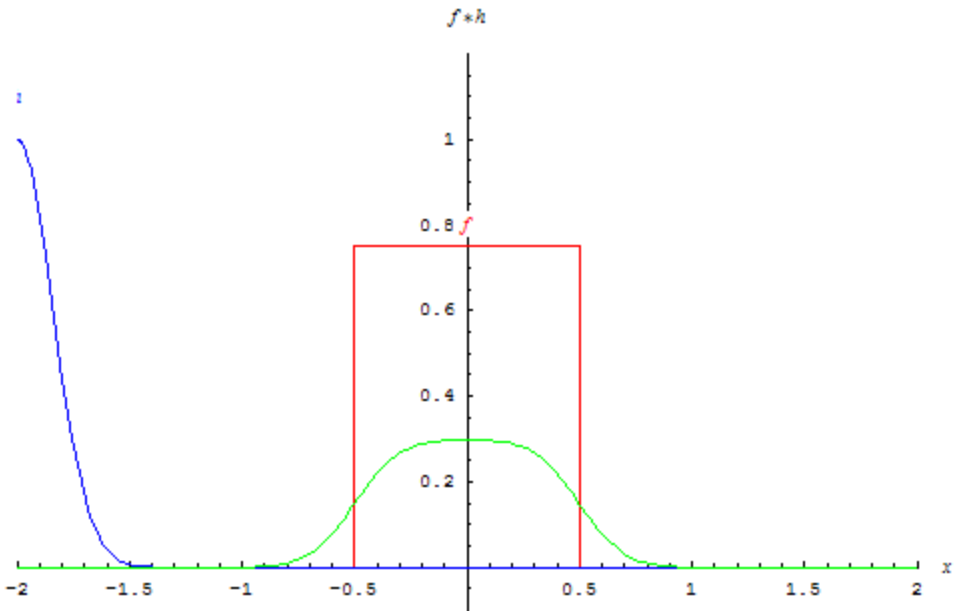
It is probably best to illustrate convolution with some examples. In each example, the blue line represents the function $h(x - x')$, the red line the function $f(x)$ and the green line is the convolution. In the animation; follow the vertical green line that is the point where the convolution is being evaluated. Its value is the area under the product of the two curves at that point.



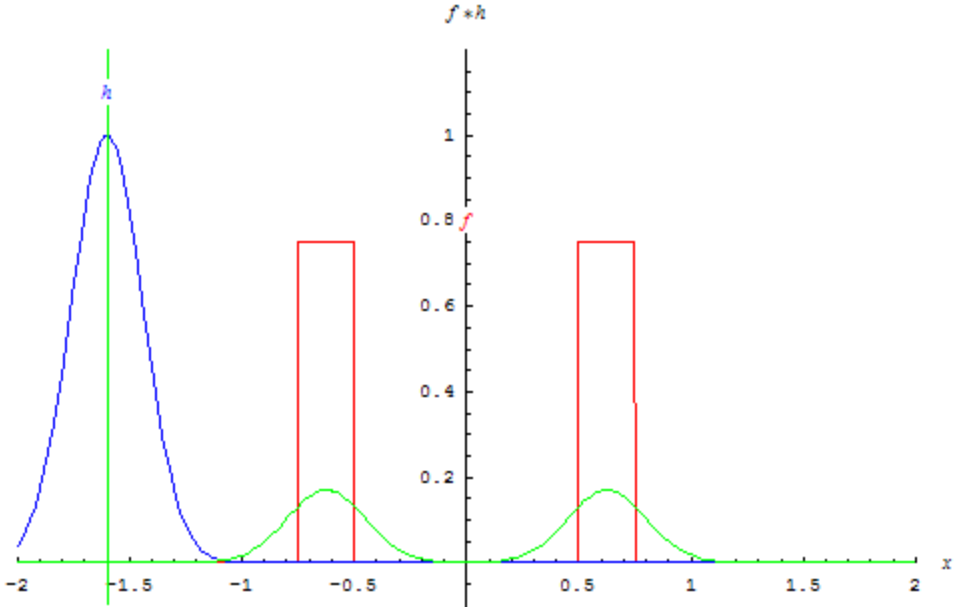
It might be easier to picture what is going on if we capture a couple of frames.



Here is a slightly more complicated example



Finally it is interesting to note what happens when we spread out a few functions, that is in this case, f is a step function in a couple of places.



The convolution theorem states that if

$$G(k) = F\{g(x)\}$$

$$F(k) = F\{f(x)\}$$

and

$$H(k) = F\{h(x)\}$$

and if

$$g(x) = f \otimes h$$

then

$$G(k) = F(k)H(k).$$

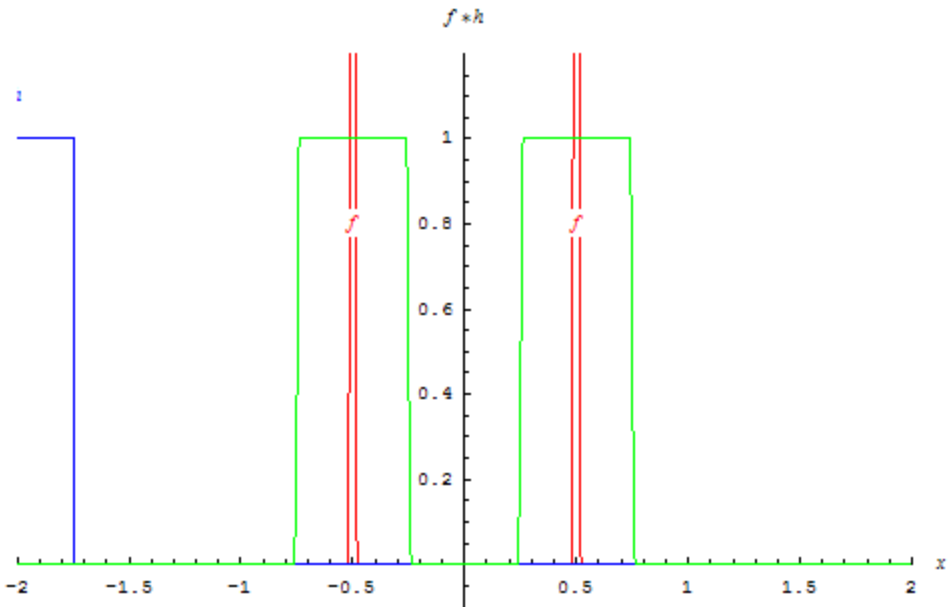
We can easily show this

$$\begin{aligned} G(k) &= \int_{-\infty}^{\infty} g(x')e^{ikx'}\mathrm{d}x' \\ &= \int_{-\infty}^{\infty} e^{ikx'} \int_{-\infty}^{\infty} f(x)h(x'-x)\mathrm{d}x\mathrm{d}x' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x'-x)e^{ikx'}\mathrm{d}x' f(x)\mathrm{d}x \end{aligned}$$

now set $w = x' - x$ then $x' = w + x$

$$\begin{aligned} G(k) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(w)e^{ikw}\mathrm{d}w e^{ikx}\mathrm{d}x \\ &= \int_{-\infty}^{\infty} h(w)e^{ikw}\mathrm{d}w \int_{-\infty}^{\infty} f(x)e^{ikx}\mathrm{d}x \\ &= H(k)F(k) \end{aligned}$$

Now say we want to consider the case of two long slits with width a. This can be described by the convolution of one slit with two delta functions. Unfortunately it is not possible to animate this since the delta function is infinitely narrow. However an extremely narrow Gaussian is a good approximation to the Dirac delta function and I have used that for the animation below.



So two slits of a finite width can be described by the convolution of two delta functions and rectangular aperture function. Then the Fraunhofer diffraction pattern is just the product of the two Fourier transforms.

To sumarize: Fraunhofer diffraction patterns are the Fourier transform of the aperture function. The Fourier transform of the convolution of functions is the product of the Fourier transforms of the individual functions. each of our complex diffraction cases could be considered the convolution of simpler cases, hence the resulting patterns were the products of those simpler cases.

An Array of Apertures

Consider an array of rectangular slits of widths a and b with spacing d in the y direction, illuminated by a plane wave. We want to derive the Fraunhofer diffracted field $E(k_x, k_y)$. The simplest way to do this problem is to use Fourier optics: The array is a convolution of a single rectangular slit with an array of Dirac delta functions.

That is the aperture function can be written

$$A = A_0(f \otimes h)$$

where the symbol $\otimes$ represents convolution and

$$\begin{aligned} f = 1 \qquad & when \quad |x| \leq a/2 \quad and \quad |y| \leq b/2 \\ = 0 \qquad & otherwise \end{aligned}$$

$$h = \sum_{n_y=0}^{N-1} \delta(y - n_y d).$$

We now have that

$$E = \frac{\mathcal{E}_A}{R} e^{-i(kR-\omega t)} \mathcal{F}\{f\}\mathcal{F}\{h\}$$

where R is the distance from the origin to the point of measurement. and $F\{\}$ represents a Fourier transform. Now we have

$$\mathcal{F}\{f\} = \int_{-b/2}^{b/2} \int_{-a/2}^{a/2} e^{i(k_x x' + k_y y')} \mathbb{d}x' \mathbb{d}y'$$

or rearranging,

$$\mathcal{F}\{f\} = \int_{-b/2}^{b/2} \mathbb{d}y' e^{ik_y y'} \int_{-a/2}^{a/2} \mathbb{d}x' e^{ik_x x'}.$$

These are integrals we have done a number of times already:

$$F\{f\} = bsinc(k_y b/2) a sinc(k_x a/2)$$

Next we transform the sums of the delta functions

$$F\{h\} = \sum_{n_y=0}^{N-1} \int_{-\infty}^{\infty} e^{ik_y y} \delta(y - n_y d) dy,$$

which upon integration becomes

$$F\{h\} = \sum_{n_y=0}^{N-1} e^{ik_y n_y d}.$$

We have also performed this sum before and it gives

$$F\{h\} = e^{ik_y (N-1)d/2} \frac{\sin\left(N k_y d/2\right)}{\sin\left(k_y d/2\right)}.$$

So we have

$$E = \frac{\varepsilon_A}{R} e^{-i(kR - \omega t)} e^{ik_y (N-1)d/2} \frac{\sin\left(N k_y d/2\right)}{\sin\left(k_y d/2\right)} bsinc(k_y b/2) a sinc(k_x a/2)$$

or if we define

$$kR_c = kR - (N-1)dk_y/2$$

$$E = \frac{\varepsilon_A ab}{R} e^{-i(kR_c - \omega t)} \frac{\sin\left(N k_y d/2\right)}{\sin\left(k_y d/2\right)} sinc(k_y b/2) sinc(k_x a/2).$$